

**Forsvarets høgskole**

**Våren 2012**

**Masteroppgave**

**Seleksjon til befalsskolen**

*En evaluering av seleksjonssystemets prediktive validitet*

**Trond Joar Kjenstadbakk**



## Forord

Arbeidet med denne studien har gitt meg en unik mulighet til å kunne fordype meg i de metodene vi benytter oss av ved Forsvarets Opptak og Seleksjon til befalskolen. Det er mye forskning som er utført innen denne delen av psykologien, men allikevel opplever jeg at det samtidig er utrolig mange ledere som uttaler seg skråsikkert om hva som fungerer uten å ha vitenskapelig dekning for det. Denne studien håper jeg vil være et godt hjelpemiddel for å kunne utvikle og bedre de seleksjonsmetodene som benyttes i et av de kanskje viktigste seleksjonsprosessene inn til Forsvaret.

I løpet av studien har jeg fått et betydelig mer åpent og ydmykt forhold til forskning, og statistikk spesielt, som et hjelpemiddel og verktøy i evaluerings- og utviklingsprosesser. Generelt tror jeg vi alt for ofte benytter vår magefølelse i stedet for å støtte oss på forskning når det gjelder vurderinger på hva som fungerer eller ikke. Den gylne middelvei er kanskje her også den rette vei.

Jeg vil rette en stor takk til min veileder, Monica Martinussen ved Universitetet i Tromsø/Forsvarets Høgskole, for en uvurderlig støtte og veiledning i løpet av hele prosessen. Med sin kunnskap og erfaring innen seleksjonsfaget har hun uten tvil løftet oppgaven til det den er. I tillegg vil jeg takke Sjefpsykolog i Forsvaret Ole Christian Lang-Ree for all støtte, og for å inkludere meg inn i sin seksjon. Det har vært en inspirasjon og glede å få ta del i. Det sitter i tillegg to meget hjelpsomme damer ved biblioteket ved Forsvarets Høgskole som ikke kan fullroses godt nok. Takk, Silje og Nina! Videre vil jeg rette en stor takk til min sparringspartner Førsteamanuensis Eivind Åsrum Skille ved Høgskolen i Hedmark for meget nyttige diskusjoner rundt studien. Det er i tillegg flere som har bidratt og hjulpet til i prosessen både fra Vernepliktsverket og Hærens befalsskole. Takk for all støtte og hjelp.

Trond Joar Kjenstadbakk

Forsvarets Stabsskole 23. mai 2012

## Abstract

This study seeks to examine the predictors used by the Norwegian Armed Forces to select candidates to Non Commissioned Officer training. By using data from the Joint Selection Centre and results after one year training, the study uses correlation and regression analysis to examine the predictive validity of the different predictors. The predictors used in the study are rank from High School, general mental ability (GMA), prognoses assessed in an interview, physical fitness tests and assessment center. These predictors were compared with three criteria, academic performance, leadership performance and total score.

Rank in school and GMA show a good predictive validity with both academic performance and total score. Academic prognosis from the interview shows a good predictive validity with academic performance and total score as well, but the leadership prognosis shows a rather weak predictive validity with leadership performance. Physical test shows low predictive validity with all the criteria. Assessment centre however shows a good predictive validity with especially leadership performance, but also with total score. Most of total variance of academic performance is explained by rank in school and GMA. Most of total variance of leadership performance is explained by the assessment centre. When it comes to total score most of the total variance is explained by rank in school and GMA, but especially assessment centre brings in additional explanation with statistical significance.

## Sammendrag

Denne studien hadde til hensikt å finne ut om de metodene som benyttes ved Forsvarets Opptak og Seleksjon til befalsskolen (FOS BS) faktisk virker etter hensikten. Klarer man å predikere fremtidige prestasjoner med de metodene som benyttes? For å finne ut av det ble det tatt utgangspunkt i et kull fra Hærens befalsskole (HBS). Data fra FOS BS ble sammenlignet med resultater som kandidatene presterte etter ett års skole. De prediktorene som ble testet var skolepoeng fra videregående, alminnelig evnenivå (AE), skoleprognose og lederprognose fra intervju, fysisk test og offisersvurdering FOS (vurderingssenter). Kriteriene som de ble målt opp mot var snittkarakterer på teorifag som måler akademisk prestasjon, offisersvurdering HBS som måler lederprestasjoner og til slutt total prestasjon i form av hovedkarakter.

Studien så først på korrelasjoner mellom de ulike variablene over. Både inter-korrelasjon og korrelasjon mellom prediktor og kriterium ble vurdert. Til slutt ble det gjennomført en multipl hierarkisk regresjonsanalyse som skulle finne ut av total forklart varians for de ulike kriteriene, samt hvor mye prediktorene samlet predikerer fremtidig resultat.

Skolepoeng fra videregående og AE kommer ut med en relativt høy prediktiv validitet for både akademiske prestasjoner og hovedkarakter. Det samme gjelder for skoleprognosen som blir fastsatt gjennom et intervju. Den er ment å predikere akademisk prestasjon, og den kommer ut med gode resultater. Den har en høy korrelasjon med akademisk og total prestasjon, men det er mer usikkert på hvor mye ekstra den tilfører i forklaring av total varians. Lederprognosen som også settes i løpet av intervjuet kommer ut med en lavere korrelasjon med lederprestasjon. Den settes på bakgrunn av et relativt beskjedent grunnlag og ser ut til å være noe påvirket av skoleprognosen som settes i samme intervju. Lederprognosen som den fremgår i denne studien kommer derfor ut med en relativt lav prediktiv validitet. Offisersvurdering FOS som etableres på bakgrunn av feltøvelsen, eller et vurderingssenter, viser til en relativt sterk korrelasjon med leder- og total prestasjon. Den fremstår også som en unik bidragsyter i forklaring av total varians for både leder- og total prestasjon, med statistisk signifikans. Fysisk test viser lav eller ingen korrelasjon med noen av de kriterier som ble benyttet i oppgaven.

Alt i alt synes de fleste prediktorer som benyttes ved FOS BS til å ha en relativt høy prediktiv validitet mot de valgte kriterier. Allikevel kan det være forbedringspotensiale både for de prediktorer som benyttes, men også et spørsmål om endring av vekting i opptaket, samt tilføring av eventuelt andre prediktorer.

## Innholdsfortegnelse

<b>Forord .....</b>	<b>3</b>
<b>Abstract .....</b>	<b>4</b>
<b>Sammendrag .....</b>	<b>5</b>
<b>1 Innledning .....</b>	<b>7</b>
1.1 GENERELT .....	7
1.2 FELLES OPPTAK OG SELEKSJON TIL BEFALSSKOLEN .....	7
1.3 HVORFOR ER SELEKSJON VIKTIG Å FORSKE PÅ? .....	9
1.4 PROBLEMSTILLING .....	10
<b>2 Forskning på seleksjon .....</b>	<b>11</b>
2.1 VALIDERINGSSTUDIER .....	11
2.2 GENERELT OM FORSKNING PÅ SELEKSJON .....	13
2.3 SKOLERESULTAT FRA VIDEREgåENDE SKOLE .....	14
2.4 ALMINNELIG EVNENIVÅ .....	15
2.5 FYSISK TEST .....	16
2.6 INTERVJU .....	16
2.7 VURDERINGSSENTER .....	17
2.8 FORVENTNINGER TIL FUNNENE .....	17
<b>3 Bruk av prediktorer ved FOS BS .....</b>	<b>19</b>
3.1 BESKRIVELSE AV FELLES OPPTAK OG SELEKSJON BEFALSSKOLEN .....	19
3.2 OPPTAKSUK .....	20
3.3 FELTØVELSE .....	23
3.4 HOVEDOPPTAK .....	23
<b>4 Metode .....</b>	<b>24</b>
4.1 DELTAGERE .....	24
4.2 PREDIKTORER .....	24
4.3 KRITERIER .....	25
4.4 PROSEDYRE .....	26
4.5 STATISTISKE ANALYSER .....	26
<b>5 Resultater .....</b>	<b>28</b>
5.1 KORRELASJONSANALYSE .....	28
5.2 MULTIPPEL HIERARKISK REGRESJONSANALYSE .....	28
<b>6 Diskusjon .....</b>	<b>31</b>
6.1 KORRELASJONER .....	31
6.1.1 Skolepoeng videregående skole og skoleprognose .....	31
6.1.2 Alminnelig evnenivå .....	33
6.1.3 Lederprognose .....	34
6.1.4 Offisersvurdering FOS .....	36
6.1.5 Fysisk test FOS .....	38
6.1.6 Reliabilitet og validitet for kriterier .....	39
6.2 REGRESJONSANALYSE .....	40
6.2.1 Akademisk prestasjon .....	40
6.2.2 Lederpotensiale .....	41
6.2.3 Total prestasjon .....	42
6.3 STYRKER OG SVAKHETER VED STUDIEN .....	43
<b>7 Konklusjon .....</b>	<b>45</b>
7.1 AKADEMISKE IMPLIKASJONER .....	46
7.2 PRAKTISKE IMPLIKASJONER .....	47
<b>Referanser .....</b>	<b>49</b>

## 1 Innledning

### 1.1 Generelt

*Plutselig støter en lagfører og en soldat i fremskutte posisjoner på en fiendtlig styrke på 10-15 personer, som forbereder et angrep. De havner i en intens skuddveksling på et par hundre meters hold. De to nordmennene skyter seg tomme for ammunisjon med automatgeværene. Men lagføreren – en fenrik – holder hodet kaldt. De klarer å få et maskingevær i stilling og fortsetter kampen. Fienden får panikk og trekker seg ut. Lagføreren gir så nøyaktig informasjon til en forsterkningsstyrke som kan gå til motangrep med tyngre våpen.*

(Bentzrød, 2011)<sup>1</sup>

Relativt mange i samfunnet har et forhold til seleksjon. Seleksjon gjennomføres i de fleste organisasjoner i en eller annen form. Målsettingen med seleksjonen er å velge ut de søkerne som har størst forutsetning for å klare utdanningen eller utføre jobben godt (Farr & Tippins, 2010, s. 1). Situasjonen i eksemplet fra Afghanistan kan virke ekstrem, men den er høyst reell. Det er viktig at Forsvaret klarer å selektene nettopp de personer som kan takle slike situasjoner. Det finnes en rekke metoder som kan anvendes i seleksjonsprosessen. Det kan være bruk av psykologiske tester, intervju eller at søkerne observeres mens de gjennomfører ulike oppgaver (Braun, Wiegand, & Aschenbrenner, 1991; Schmidt & Hunter, 1998). Et viktig spørsmål i denne sammenhengen er om de ulike metodene faktisk fungerer slik de skal, og gjør seleksjonen mer presis enn tilfeldig utvelgelse eller selv-seleksjon. Det studeres vanligvis ved at man undersøker sammenhengen mellom resultatene fra seleksjonen (f.eks. testresultater eller prognoser) og senere jobb- eller skolerresultater (Farr & Tippins, 2010; Martinussen & Hunter, 2008). Det er vanlig å stille kvalitetskrav til metoder som skal anvendes til å ta viktige beslutninger om personer som for eksempel tilgang til utdanning eller jobb (Bartram, 2001). Det er viktig at metodene er både reliable, det vil si måler konsistent og korrekt, og at man i tillegg kan dokumentere at de faktisk er egnet til å si noe om personens potensiale og framtidige arbeidsprestasjoner. Er de vurderinger som utføres i forbindelse med en seleksjonsprosess av en slik karakter at de predikerer eller forutsier fremtidige prestasjoner? For å kunne dokumentere slike forhold er det vanlig å gjennomføre empiriske undersøkelser.

### 1.2 Felles Opptak og Seleksjon til befalsskolen

Hver sommer gjennomfører alle befalsskoler i Forsvaret Felles Opptak og Seleksjon (FOS BS). Målet med denne perioden er "... å identifisere kandidater som har potensiale til å nå Forsvarets ambisjoner med hensyn til faglige, lederskapsmessige og fysiske krav" (Forsvarets skolesenter, 2007, s. 72). Opptaket består av en rekke tester og vurderinger som skal gjøre befalsskolene i stand til å selektene de best egnede til videre utdanning og karriere i Forsvaret. Denne oppgaven

<sup>1</sup> Artikkel i Aftenposten (17. mars 2011) som refererer til en rapport fra Afghanistan, og beskriver en reell situasjon.

vil ta for seg Hærens befalsskole (HBS), men det vil helt klart kunne sies å være likheter med de andre befalsskolene i Forsvaret. Forsvaret har i en årrekke gjennomført seleksjon av ungdommer som ønsker seg en militær karriere. Det inkluderer seleksjon til krigsskolene, flygere, og ulike tekniske fagskoler i tillegg til befalsskolen. Innholdet, altså hvilke formelle krav som stilles til søkerne og hvilke metoder som anvendes, har endret seg over tid både i Norge og internasjonalt (Jones, 1989, s. 78).

Et viktig grunnlag for en vellykket seleksjon er at en vet hvilke krav som søkerne stilles overfor både på kort og noe lengre sikt (Morgeson & Campion, 2000; Schmidt, Ones, & Hunter, 1992, s. 654-655). Hvilke prestasjoner, ferdigheter og evner som vil være av betydning for å gjøre en god jobb eller klare å gjennomføre en utdanning er et viktig utgangspunkt for en god seleksjonsprosess. Utarbeidelsen av en slik oversikt tar utgangspunkt i den jobben som skal utføres etter seleksjon og benevnes ofte som en jobbanalyse (Martinussen & Hunter, 2008, s. 110). Det finnes ulike metoder man kan anvende for å gjennomføre en jobbanalyse, og flere av disse baserer seg på at man anvender eksperter som beskriver sentrale arbeidsoppgaver og deretter rangerer ulike evne- og personlighetmessige egenskaper som er viktige for å kunne utføre disse arbeidsoppgavene. Deretter velger man tester og målemetoder som kartlegger de aktuelle evnene og egenskapene. I tillegg er det imidlertid viktig å sjekke empirisk at metodene faktisk predikerer framtidig jobb- og skoleprestasjoner. Selv om en jobbanalyse kan sies å være et viktig trinn på veien i forhold til å velge ut gode tester og seleksjonsmetoder vil denne studien i liten grad berøre denne delen av prosessen, men i stedet omhandle en evaluering av metodenes prediktive validitet. Prediktiv validitet – viser til hvor godt de ulike seleksjonsmetodene eller prognosene faktisk predikerer framtidige jobb- eller skoleprestasjoner som den enkelte kandidat leverer (Hardison, Sims, & Wong, 2010, s. 9).

Før selve opptaket starter er det tilgjengelig to viktige resultater i form av skolepoeng fra videregående og alminnelig evnenivå (AE). Ved FOS blir det som nevnt over gjennomført ulike tester og vurderinger av de aktuelle kandidatene. Resultatene fra de ulike testene og vurderinger av den enkelte søker summeres opp i ulike prognoser som settes for den enkelte kandidat. Den ene er lederprognose – som skal si noe om potensialet for å kunne bli en dyktig leder. En annen er skoleprognose – som skal si noe om potensialet for å gjøre det bra fagmessig og læringsmessig under utdanningen. Begge prognosene settes på bakgrunn av ulik bakgrunnsinformasjon og et intervju. De to prognosene er nylig etablert. I tillegg gjennomføres fysisk test i første del av opptaket.



I siste del av opptaket gjennomføres det en feltøvelse hvor kandidatene prøves i ulike roller og får oppgaver som er relevante i forhold til den påfølgende tjenesten i Forsvaret. Denne perioden ligger nært det som innen seleksjon kalles vurderingssenter (*assessment center*). Resultatet av denne perioden oppsummeres i en offisersvurdering. En gjennomgang av selve opptaket på FOS BS vil bli behandlet nærmere i et senere kapittel.

### 1.3 Hvorfor er seleksjon viktig å forske på?

I og med at offisersyrket kan sies å være en profesjon som hovedsaklig rekrutterer hovedtyngden av sine medarbeidere gjennom den samme trakten, vil seleksjonen til befalsskolene være meget viktig (FFOD, 2007, s. 157).

For det første er det forbundet store kostnader med å selektare feil personell. Utdanning og kompetansebygging er kostbart og kandidater som faller fra i løpet av utdanningen representerer en utgift uten at Forsvaret får noe igjen i form av ferdig utdannende offiserer. I et virksomhetsstyringsperspektiv ønsker organisasjoner å holde kostnadsbildet så lavt som mulig, og i så måte vil en kvalitativ god seleksjon ikke bare være en vel anvendt ressurs, men faktisk kunne sies å være kostnadsbesparende (Jones, 1989, s. 73). Personer som etter endt utdanning produserer lite eller bidrar lite til organisasjonens overordnede målsetning, vil representere et tap i forhold til om man hadde selektert mer produktive arbeidstakere. Det er mulig å gjennomføre ulike beregninger for å kartlegge hvor mye man tjener på seleksjonen basert på vurderinger av følgende parametere; baserate (hvor mange i søkergruppen ville ha vært kvalifisert), seleksjonsratio (hvor mange som skal velges ut av søkergruppen) og hvor god den prediktive validiteten er til den aktuelle metoden, f.eks. en konkret test (Martinussen & Hunter, 2008, s. 131-133).

For det andre har utviklingen av Forsvaret, fra et invasjonforsvar med utdanning til mobiliseringshær til et innsatsforsvar med daglig utnyttelse av vårt befalskorps, sies å ha medført et behov for en enda strammere seleksjon. Dette innebærer at det er et større behov for at flere av de som utdannes gjennom befalsskolen også fortsetter i tjeneste som befal (Forsvarsdepartementet, 2008, s. 111).

Et tredje perspektiv er at mange av de som selekteres inn i Forsvaret skal ut i skarpe oppdrag i internasjonale operasjoner og lede mannskaper under svært krevende situasjoner. En rekke aktiviteter som utføres i Forsvaret er forbundet med risiko, og som offiserer skal de kunne fremstå som dyktige ledere under et slikt operasjonsmiljø. Konsekvensene av å mislykkes innen seleksjon kan føre til situasjoner som i ytterste fall resulterer i tap av menneskeliv. Større fokus

på realistisk trening hjemme, samt en betydelig leveranse av styrker til skarpe internasjonale oppdrag krever i større grad enn tidligere en vellykket seleksjon.

Med andre ord er det minst like viktig i dag som tidligere at vi sikrer en god seleksjonsprosess. Det kan være fristende å holde seg til de verktøy som organisasjonen er kjent med og som "alltid" har fungert. På en annen side kan slike ukritiske tanker føre til dårlig seleksjon som ikke møter de behovene organisasjonen faktisk har. I praksis betyr dette at en jevnlig bør sjekke at de metodene en benytter fungerer etter hensikten (Kahneman, 2011). Videre underbygges denne problematikken i en artikkel som tar for seg forskjellen mellom forskningsbasert validitet for ulike metoder sammenlignet med vurdert eller estimert validitet av konsulenter innen seleksjon i New Zealand (Dakin & Armstrong, 1989, s. 188). Den viser til dels overraskende store forskjeller mellom forskningsbasert validitet, og det som konsulenter innen seleksjon tror eller mener fungerer best som metode for å forutsi fremtidige prestasjoner.

#### 1.4 Problemstilling

Når det gjelder skole- og lederprognosen har disse kun vært benyttet av Hæren i to år. Ved FOS BS sommeren 2011 ble de imidlertid innført for alle befalsskolene. Når denne oppgaven skrives er det kun tilgjengelig resultat for ett kull fra HBS som ble tatt opp sommeren 2010. Det finnes dermed et begrenset datautvalg som både har prediktorer og kriterier tilgjengelig. Det er imidlertid ikke gjennomført noen studier på om de nye prognosene faktisk er relatert til senere skolerresultater. En av motivasjonene for denne oppgaven vil derfor være å undersøke om det er noen sammenheng mellom disse prognosene og de faktiske skolerresultater etter endt skoleår. Disse to prognosene utgjør bare en del av det som vektlegges under selve opptaket. Det er også av interesse å se på hvor godt de andre metodene som benyttes predikerer fremtidige resultater.

Mer presist formulert vil formålet med denne studien være å estimere den prediktive validiteten til enkelte sentrale prediktorer som anvendes ved seleksjon til befalsskolen. I tillegg er det av interesse å undersøke hvor godt de ulike prediktorer samlet sett predikerer prestasjoner etter endt skoleår.

#### Problemstillinger:

*Hvilken sammenheng er det mellom de prediktorer som nyttes ved Felles Opptak og Seleksjon til Hærens befalsskole og skolerresultat etter endt utdanning?*

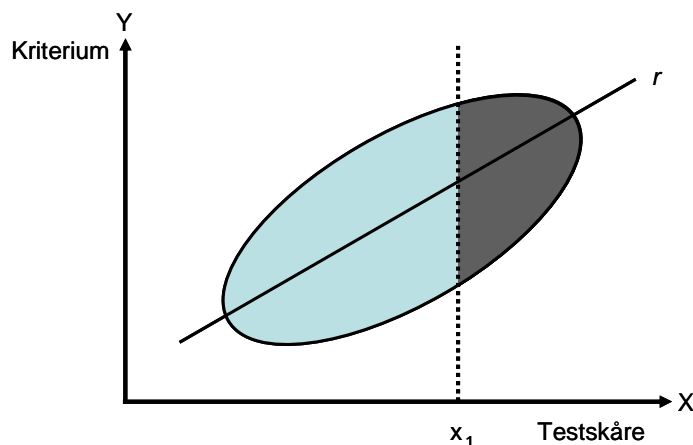
*Hvor mye av den totale variansen for skolerresultat forklares av de ulike prediktorer samlet sett?*

*Hvor mye av den totale variansen for skolerresultat forklares av de prediktorer som eksisterer før opptaket, og hvor mye tilføres av de som etableres under selve opptaket?*

## 2 Forskning på seleksjon

### 2.1 Valideringsstudier

For å undersøke om de prediktorer som nyttes i forbindelse med seleksjon faktisk fungerer i henhold til intensjonene, gjennomføres normalt en lokal valideringsstudie (Martinussen & Hunter, 2008, s. 117-118). Prediktorer er benevnelsen på de ulike metodene som benyttes ved seleksjon (Martinussen & Hunter, 2008, s. 113). Ved en lokal valideringsstudie beregnes korrelasjonene mellom de ulike prediktorer som nyttes i seleksjonsøyemed opp mot ulike vurderinger av senere prestasjoner. Disse prestasjonene benevnes ofte som kriterier i forbindelse med seleksjon (Martinussen & Hunter, 2008, s. 113). Et problem knyttet til denne typen undersøkelser er at jo strengere seleksjonen er, desto mindre blir gruppen som det er mulig å inkludere i slike studier. Kandidatene som faller fra underveis i prosessen mangler data i form av skolerresultat, og kan derfor ikke studeres på samme måte som den utvalgte gruppen. Gjennom seleksjon blir de med høyest skåre på de ulike testene valgt, noe som medfører redusert spredning på prediktor - variablene. Dette fører igjen til at de beregnede korrelasjonene blir lavere enn om man hadde hatt muligheten til å studere hele søkergruppen (Martinussen & Hunter, 2008, s. 119-120). Utfordringen i forbindelse med en lokal valideringsstudie blir dermed at utvalget etter seleksjon er begrenset, og således gir noe skjevt utvalg ved at kontrollgruppen er sendt hjem. Dette metodiske problemet omtales ofte som "*restriction of range*" (Martinussen & Hunter, 2008, s. 119-120). Det er mulig å foreta korreksjoner for å få et bedre anslag på den egentlige prediktive validiteten. Imidlertid er det svært komplisert i situasjoner med mange prediktorer og en trinnvis seleksjon. Derfor blir ikke slike korreksjoner gjennomført i denne oppgaven. Det som imidlertid er sannsynlig er at den prediktive validiteten til de prediktorer som analyseres i denne oppgaven hadde blitt enda høyere dersom det hadde vært mulig å gjennomføre slike korreksjoner (Borack, 1994).



Figur 1.1 Modell av "restriction of range"-fenomenet

Problemet med *restriction of range* kan illustreres ved figur 1.1, som viser sammenhengen mellom testresultater og kritereskårer for hele gruppen og for den selekterte gruppen.

Sammenhengen, eller korrelasjonen som beregnes for den gruppen til høyre (mørkt felt) blir da betraktelig svakere, enn om hele gruppen hadde blitt undersøkt. Dette kan sees på som et "missing data" problem, der frafallet er systematisk i stedet for tilfeldig.

En annen utfordring med lokal valideringsstudie er ofte lav  $n$ . Utvalget som en står igjen med etter seleksjon kan være meget begrenset, i tillegg til at mange organisasjoner ansetter for få personer om gangen slik at utvalget også av denne grunn blir lite. I opptaket til befalsskolen er fortsatt utvalget på ca 200 etter at opptaket er ferdig, men ser man på tradisjonell jobbutvelgelse er som oftest utvalget betraktelig mindre. Det fører til lav statistisk styrke (*statistical power*) som innebærer at det kan bli vanskelig å avdekke en signifikant effekt, selv om en sann effekt eksisterer, altså at det er en positiv korrelasjon mellom prediktor og kriteriet (Park, 2010, s. 7). I tillegg vil valg av kriterium være viktig for en slik valideringsstudie. Hvilke prestasjoner skal seleksjonsmetodene i utvalget testes eller måles mot, og hvor mange kriterier bør en velge ut? Det kan være alt fra skolerresultater til vurderinger utført av ledere, mens i andre sammenhenger kan det være mer objektive resultater i form av produksjonsmål. Spørsmålet er ofte hvor egnet og pålitelig (reliable) kriteriene som velges ut er. Den prediktive validiteten estimeres ofte i form av en korrelasjon mellom prediktor og kriterium, og dersom det anvendes flere prediktorer kan multipl regressjonsanalyse anvendes for å beregne hvor mye varians i kriteriet som er forklart ved hjelp av en lineær kombinasjon av de aktuelle prediktorer (Ringdal, 2009, s. 361-387).

Et alternativ til å gjennomføre lokale valideringsstudier er meta-analyser. Meta-analyser er en fellesbetegnelse på at man foretar en kvantitativ oppsummering av funn (effekter) fra

tidligere undersøkelser (Jones, 1989, s. 65). De undersøkelsene som sammenstilles vil ofte være lokale valideringsstudier, og for hver studie registreres først den aktuelle korrelasjonskoeffisienten og så utvalgets størrelse ( $N$ ). Gjennom en meta-analyse beregnes en gjennomsnittlig vektet korrelasjon basert på alle de undersøkelsene som tas med. Den gjennomsnittlige korrelasjon vektet i forhold til utvalgenes størrelse slik at en korrelasjonskoeffisient beregnet for et stort utvalg får større betydning enn en korrelasjon beregnet for et lite utvalg. Det vil i tillegg være mulig å studere variasjon mellom studier og om det er slik at validiteten lar seg generalisere over mange ulike situasjoner og utvalg. Det finnes flere ulike statistiske teknikker for å gjennomføre en meta-analyse, men i forhold til valideringsstudier er det mest vanlig å anvende Hunter og Schmidts metode (2003), der man også korrigerer for ulike feilkilder som *restriction of range* og manglende reliabilitet i kriteriet.

## 2.2 Generelt om forskning på seleksjon

Det finnes i dag omfattende forskning som evaluerer ulike seleksjonsmetoder, og hvor godt de klarer å forutsi fremtidige resultat. Seleksjon av kandidater som har liten eller ingen erfaring innen det de skal selekteres til har også blitt viet stor oppmerksomhet. Dette er som oftest tilfelle når det gjelder seleksjon til befalsskole i Forsvaret. Forskningen innen seleksjon spenner fra store kunnskapsoppsummeringer som har sammenlignet flere studier fra flere årtier og omhandler flere metoder, til mer spesifikke studier som tar for seg enkelte metoder (Hunter & Hunter, 1984; Russell & Domm, 1995; Schmidt & Hunter, 1998).

Militær seleksjon har lange tradisjoner. Imidlertid meldte behovet seg etter hvert for å ta i bruk mer objektive metoder til denne utvelgelsen enn tidligere erfaring og referanser. Allerede fra første verdenskrig hadde mange vestlige land tatt i bruk både vurdering av medisinske krav, satt krav om tidligere utdanning så vel som tatt i bruk psykologiske tester (Jones, 1989, s. 63-65). I Norge har psykologiske tester vært i bruk til dette formålet i tiden etter andre verdenskrig (Hansen, 2007). En stor andel av testutviklingen innenfor militær seleksjon og senere validering har skjedd på amerikanske utvalg (Knapp & Tremble, 2007). En av de mest fremtredende studiene innen amerikansk militær forskning på seleksjon er utvilsomt *Project A*, som så på hele det amerikanske Forsvarets metoder i et tolv års perspektiv (Campbell & Knapp, 2010). Det er også gjennomført enkelte europeiske studier, men utvalgene er jevnt over mindre (Martinussen & Torjussen, 2004). Forskning på seleksjon i Europa har gått i bølgedaler, fra en aktiv og solid forskning fra starten på 1900-tallet frem til andre verdenskrig. Frem til 1970 var det en periode karakterisert av stagnasjon før det igjen tok seg opp og i dag fremstår som et meget viktig forskningsfelt (Salgado, Anderson, & Hulsheger, 2010). Det er gjennomført en rekke studier

både på sivile og militære søkere for å anslå metodenes prediktive validitet (Schmidt & Hunter, 1998). Generelt har evnetester og vurderingssentre (*assessment centre*) hatt god prediktiv validitet, mens ustrukturerte intervjuer og personlighetstester har hatt lavere prediktiv validitet (Schmidt & Hunter, 1998). Personlighetstester har vært brukt i mindre grad til seleksjon, og i perioder har man tvilt på testenes prediktive validitet (Guion & Gottier, 1965). Det kan skyldes diskusjoner rundt struktur av personlighetsvariabler, og at tidligere forskning har hatt utfordringer rundt å knytte personlighetstester opp mot senere jobbprestasjoner (Hough & Dilchert, 2010, s. 299-301). Allikevel fremhever studier at personlighetstester også har en nyttig funksjon i forbindelse med seleksjon (Hystad, Eid, Laberg, & Bartone, 2011, s. 387-388), men at den prediktive validiteten har vært noe lavere enn evnetester (Martinussen & Torjussen, 2004; Schmidt & Hunter, 1998).

Etter å ha redegjort for en del generell forskning innen seleksjonsfaget, følger nå en mer spisset gjennomgang av forskning for de prediktorer som nyttes ved opptaket til befalsskolen.

### 2.3 Skoleresultat fra videregående skole

Skoleresultater fra videregående skole er en etablert prediktor innen utvelgelse til offiserstrening. Spesielt i forbindelse med utvelgelse til akademisk utdanning innen Forsvaret som krigsskole, men den nyttes også i forbindelse med opptak til befalsskolen. Denne prediktoren eksisterer allerede før selve opptaket starter, og benyttes blant annet til utvelgelse av de kandidater som skal kalles inn til selve opptaket. Skoleresultat som prediktor ble spesielt studert i en analyse for den amerikanske marinen, og fremstod med en relativt høy korrelasjon med både fremtidig akademiske ( $r = .44$ ) og totale prestasjoner ( $r = .35$ ) for en gruppe søkere til *Naval Academy* (tilsvarende sjøkrigsskolen i Norge) (Alf, Neumann, & Mattson, 1988). De aller fleste som søker befalsskolen har begrenset erfaring fra annet enn videregående skole, og på den måten blir den en viktig prediktor.

Det er også utført undersøkelser på sammenhengen mellom referanser fra videregående opp mot ulike kriterier. Referanser henviser til en skriftlig anbefaling fra sentrale lærere i den videregående skolen. De gir en vurdering av potensiale til hver enkelt elev innen gitte områder som kommunikasjon, personlig fremtreden, mellommenneskelige relasjoner og lederskap. Samme studie som over så på sammenhengen mellom slike referanser eller anbefalinger fra overordnede ved *High School*, og gitte kriterier. De viser ingen sammenheng med verken akademisk eller total prestasjon. Det er noe overraskende at det ikke vises til sterkere sammenheng mellom anbefalinger og akademisk prestasjon, fordi en burde kunne forvente at

overordnede fra skolen hadde tilgang til skoleresultat for å kunne forutsi akademisk prestasjon (Alf et al., 1988).

I en annen studie undersøkes flere prediktorer, blant annet skolepoeng fra videregående (Jones, 1989, s. 67). Den viser også til at skoleresultat fra videregående er en vel etablert prediktor inn til offiserstrening, og at den i flere studier predikerer det aktuelle kriteriet. På en annen side bør man ha et reflektert forhold til hvor stor betydning og vektning denne prediktoren bør ha. I det ligger det at offisersyrket har flere dimensjoner enn akademiske evner, som praktiske ferdigheter og lederegenskaper.

## 2.4 Alminnelig evnenivå

Alminnelig evnenivå er en annen sentral prediktor som også er kartlagt før selve opptaket starter. Den, i tillegg til skoleresultat fra videregående, benyttes som bakgrunn for å velge ut de som skal kalles inn til selve opptaket. En studie gjennomført i det sveitsiske Forsvaret fant at bruk av vurderingssenter i tillegg til AE predikerte både akademiske og militærfaglige prestasjoner (Melchers & Annen, 2010). Studien baserer seg på et stort utvalg ( $N = 844$ ) og undersøker hvor godt de metodene som benyttes i seleksjon predikerer jobbutførelse mot slutten av deres offisersutdanning. Hovedfunn fra studien var at både AE og vurderingssenter hadde høy prediktiv validitet, og at de utfylte hverandre på en slik måte at ikke en av metodene ville være tilstrekkelig. Kriteriene på jobbprestasjoner i denne studien var akademisk og militærfaglig prestasjoner målt mot slutten av deres utdanningsperiode. Tester som målte AE hadde en signifikant prediktiv validitet  $r = .26$  og  $r = .21$  opp mot akademisk og militærfaglig prestasjoner (Melchers & Annen, 2010).

I tillegg viser meta-analyser at AE predikerer relativt godt innenfor flere arbeidsområder (Schmidt & Hunter, 1998), og at den gjennomsnittlige prediktive validiteten øker med kompleksiteten i jobben (Salgado et al., 2003). Når det gjelder studier på AE er det også gjennomført egne europeiske undersøkelser som ser på om det eventuelt er forskjeller i resultat mellom USA og Europa (Hulsheger, Maier, & Stumpp, 2007). Resultatet fra denne studien kan tyde på at det er marginale forskjeller.

Men hvor godt måler kognitive tester alminnelig evnenivå? Flere års erfaring med test av AE for Forsvaret i Norge, samt god dokumentasjon underbygger dens reliabilitet og validitet som testgrunnlag (Sundet, Barlaug, & Torjussen, 2004, s. 358). Alt sett under ett vil det nok være riktig å vurdere den kognitive testen av AE som gjennomføres på sesjon i Norge som rimelig pålitelig og gyldig, selv om det også her kan være ulike feilkilder.

## 2.5 Fysisk test

Fysisk test er en annen etablert prediktor ved opptak til både befalsskolen og krigsskolen. Testene har vært relativt like over flere år, og bør således være godt kjent av kandidatene som kalles inn til opptak. På den måten burde det være mulig for de fleste å bevisst trene seg opp til å bestå disse testene. Det er også her gjennomført en rekke studier rundt ulike fysiske testers prediktive validitet.

I en studie fra *US Navy* ble sammenhengen mellom fysisk test og senere jobbutførelser evaluert (Marcinik, Hyde, & Taylor, 1995). Den fysiske testen i denne studien bestod av 500 meter svømming, sit-ups, push-ups, pull-ups og 1,5 kilometer løping. Måling av jobbprestasjon bestod av et testbatteri sammensatt av ulike fysiske øvelser som var viktig for denne kategorien personell. Studien viser til at så mange som 18,5 – 25,7 % av kandidatene strøk på enkelte øvelser i testbatteriet. De antyder nødvendigheten av å utvikle flere test-batteri under seleksjonen som i enda større grad gjenspeiler senere krav i jobben. Den opprinnelige testen i studien er nesten identisk med de tester som benyttes ved FOS BS.

En annen studie har sett på ulike fysiske testers sammenheng med senere prestasjoner i det britiske forsvaret (Rayson, Holliman, & Belyavin, 2000). Studien viser til relativt store variasjoner på hvordan enkelt-tester predikerte fremtidige resultater i form av et test-batteri sammensatt av fysiske øvelser som var jobbrelatert. Eksempler på kriterie-tester var løfting av ammunisjonsbokser, bæring av jerry-kanner over en distanse og utmarsj med full utrustning.

## 2.6 Intervju

Intervju kan vel sies å være en av de mest kjente seleksjonsmetodene uavhengig av yrke. Tidligere studier på intervju som prediktor i forbindelse med ansettelser viste derimot til relativt lave sammenhenger (Pursell, Champion, & Gaylord, 1980, s. 907-908). Etter hvert som denne metoden ble videreutviklet har den derimot vist til bedre prediktiv validitet. Studier viste at strukturerte intervju kommer ut med til dels mye høyere sammenheng enn ustrukturerte. I en studie som har sett på 85 års forskning på seleksjon (Schmidt & Hunter, 1998), viser de til en gjennomsnittlig prediktiv validitet på  $r = .51$  for strukturerte og  $r = .38$  for ustrukturerte intervju. Disse funnene understøttes av en annen meta-analyse som fant sammenhenger på  $r = .62$  for strukturerte og  $r = .31$  for ustrukturerte (Wiesner & Cronshaw, 1988, s. 284). En annen meta-analyse fant at den prediktive validiteten for intervju avhenger av flere forhold (McDaniel, Whetzel, Schmidt, & Maurer, 1994). For det første spiller innholdet i intervjuet en rolle i tillegg til om det er strukturert eller ustrukturert. Samtidig varierer validiteten av om intervjuet gjennomføres av ett individ eller av et panel. Hva som til slutt kom ut med høyest prediktiv



validitet var avhengig av kriterier det ble målt opp mot. Studien viste til en prediktiv validitet på  $r = .31$  for strukturerte intervju og  $r = .23$  for ustrukturerte intervju når kriteriet var jobbutførelse. Imidlertid kom strukturerte intervju ut med  $r = .24$  og ustrukturerte intervju  $r = .25$ , når evnen til å tilegne seg nye ferdigheter ble valgt som kriterium. Intervju gjennomført av ett individ viste til generelt høyere prediktiv validitet enn intervju gjennomført av et panel.

## 2.7 Vurderingssenter

Vurderingssenter er en seleksjonsmetode hvor kandidater prøves i jobbrelevante arbeidsoppgaver, ofte i grupper (International Task Force on Assessment Center, 2000, s. 2). I en studie som tar for seg flere metoder og undersøkelser innen seleksjon kommer vurderingssenter ut som en god prediktor med relativt høy prediktiv validitet (Schmidt et al., 1992, s. 635-637). Den viser til flere rapporter hvor den prediktive validiteten varierer fra  $r = .31$  til  $r = .37$ , avhengig av valg av kriterier. Samme studie peker på at kognitive evnetester (f.eks. måling av AE) ser ut til å ha en høyere prediktiv validitet enn vurderingssenter. Samtidig viser de til at evnetester er en betraktelig billigere metode enn vurderingssenter (Hunter & Hunter, 1984; Schmidt et al., 1992).

En annen meta-analyse som ser på vurderingssenter som prediktor, og overordnede vurderinger som kriterium fant en prediktiv validitet på  $r = .28$  (Hermelin, Lievens, & Robertson, 2007). I studien fra det sveitsiske forsvaret hadde bruk av vurderingssenter en høy prediktiv validitet (Melchers & Annen, 2010). Vurderingssenter viste seg å ha en signifikant prediktiv validitet på  $r = .31$  og  $r = .32$  for henholdsvis akademisk og militærfaglig prestasjoner. Videre kom vurderingssenter ut med en prediktiv validitet på mellom  $r = .31$  og  $r = .33$  i en studie som evaluerte validiteten av seleksjon til den britiske hæren (Dobson & Williams, 1989).

Samlet sett kan det se ut som at vurderingssenter kommer ut med en relativt høy prediktiv validitet som prediktor. Noe av diskusjonen rundt denne prediktoren kan synes å fokusere på kostnaden. Vurderingssenter kan sies å være en relativt stor kostnadsdriver, spesielt når den evalueres opp mot andre metoder.

## 2.8 Forventninger til funnene

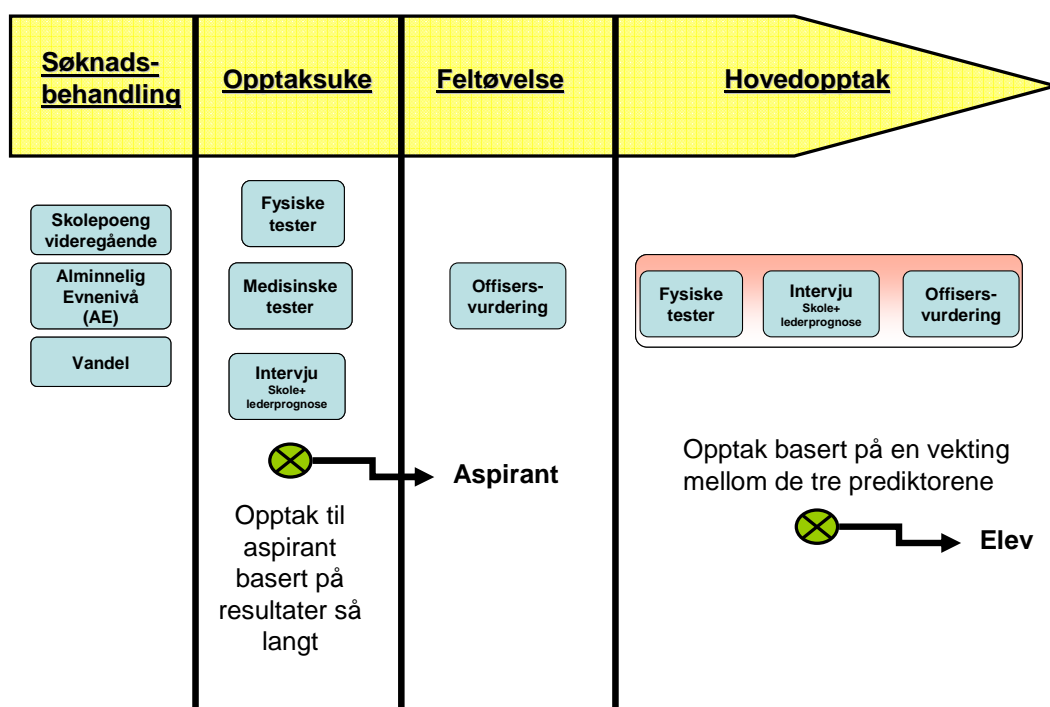
Basert på tidligere forskning innen seleksjon og prediktiv validitet bør en kunne forvente at flere av de prediktorer som nyttes under FOS BS resulterer i relativt gode korrelasjoner. Alminnelig evnenivå har både lang tradisjon i det norske Forsvaret og ellers i samfunnet. Den antas å kunne vise relativt god prediktiv validitet både mot akademiske prestasjoner og lederprestasjoner (Melchers & Annen, 2010; Salgado et al., 2003). I tillegg bør en kunne anta god prediktiv validitet for offisersvurdering FOS som baserer seg på observasjoner fra feltøvelsen (Melchers &

Annen, 2010; Schmidt et al., 1992). I det ligger at den sammenlignes med resultater for vurderingssenter. Intervju viser også til relativt god prediktiv validitet mot fremtidige prestasjoner som vist over. Derfor er det plausibelt å kunne forvente relativt sterke korrelasjoner også for de prognosene som settes på bakgrunn av intervjuet (McDaniel et al., 1994; Schmidt & Hunter, 1998).

### 3 Bruk av prediktorer ved FOS BS

#### 3.1 Beskrivelse av Felles Opptak og Seleksjon befallsskolen

Opptak og seleksjon til befallsskolen foregår i flere trinn, fra søknadsbehandling foretatt av Vernepliktsverket til opptak som elev etter endt FOS BS. Modellen skissert i figur 2.1 viser en forenklet oversikt over denne prosessen:



Figur 2-1: Seleksjonsprosessen

Alminnelig evnenivå er et gjennomsnitt av ulike evnetester som gjennomføres under sesjon før selve opptaket og er et viktig seleksjonskriterium for å bli innkalt til opptak. I tillegg benyttes skolepoeng fra videregående skole som et parameter ved innkalling til opptak.

FOS BS er felles for alle befallsskoler i Forsvaret. Tidligere ble det gjennomført opptaksuker ved de respektive befallsskolene rundt om i landet, mens nå er det samlet på ett sted under ett felles regime.

Opptaket består i hovedsak av en opptaksuke og en feltøvelse som samlet skal kunne gi nok bakgrunn for å gjennomføre en god seleksjon til befallsskolene. Underveis i opptaket er det et jevnt frafall basert på at kandidater ikke tilfredsstillir visse minimumskrav. Det gjennomføres to opptak i løpet av perioden. Det første etter endt opptaksuke, hvor kandidater selekteres videre inn i neste fase som er feltøvelsen. Det andre opptaket skjer etter endt feltøvelse og benevnes som hovedopptak. Det er her kandidatene blir tatt opp ved de respektive befallsskoler som elever

og fordelt på ulike linjer.

Alle vedtak som har med seleksjon å gjøre fattes av et skoleråd som etableres på Kjevik. Det består blant annet av alle sjefer for de respektive befalsskolene, Vernepliktsverket, samt ledelsen for FOS – organisasjonen.

### 3.2 Opptaksuke

Opptaksuken består av ulike tester som hver kandidat skal igjennom før den enkelte skole foretar sine valg av kandidater som tas opp som aspiranter og deltagere på neste fase - feltøvelsen.

Medisinske tester gjennomføres av medisinsk personell, og tar utgangspunkt i fastsatte krav for opptak som befalselev. Det er laget en minimumsprofil på helse som hver enkelt kandidat må tilfredsstille. Erfaringsvis faller det fra kandidater også her, selv om det er gjennomført en pre-seleksjon basert på resultat fra sesjon. Det kan skyldes skader og andre endringer i helseprofil som har skjedd i perioden mellom sesjon og opptak.

Fysiske tester består av styrketest, 200 meter svømming og 3000 meter løping. Styrketesten består av push ups, sit ups og hang ups. Det opereres med minimumskrav på fysiske tester for å bli med videre i opptaket. Kandidater som ligger litt under kravet, men som gjør det bra på de andre prediktorer kan gis mulighet for en re-test senere i opptaket. Avgjørelse om dette tas i skolerådet. Under opptaksuken gjennomføres det et intervju av hver enkelt kandidat. Det gjennomføres som et delvis strukturert intervju. Personer som gjennomfører intervjuet vurderes som erfarne offiserer. I tillegg er det tilgang til mer erfarne intervjuere under hele prosessen, samt at det er tilgang til psykologer i de tilfeller der det føles nødvendig. Intervjuet følger en fastlagt mal, med en viss frihet tillagt hver enkelt intervjuoffiser. Intervjuet gjennomføres for HBS sitt vedkommende av ca 20 intervjuoffiserer. Det gjennomføres et intervjuoffiserskurs av ca fire timers varighet i oppstarten av opptaksuken, i tillegg til at det er utarbeidet en egen intervjuguide<sup>2</sup>. HBS står selv for utvelgelse av hvilke offiserer som skal stå for intervjuene. Det settes av 60 minutter til hvert intervju, inkludert for- og etterarbeid.

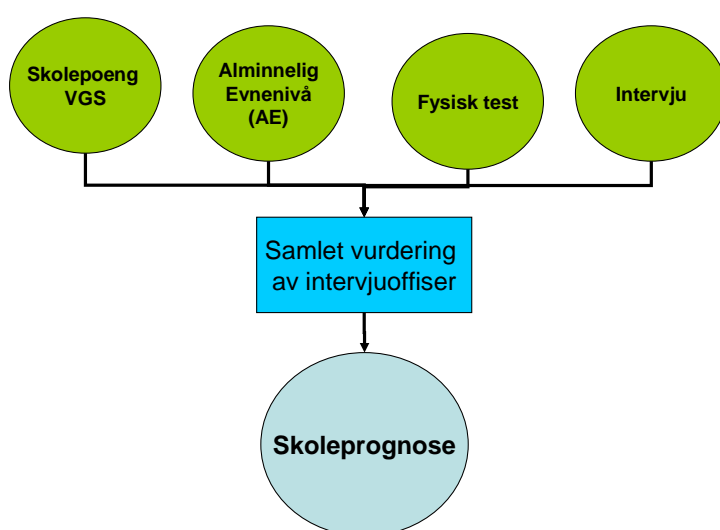
Hensikten med intervjuet er i følge intervjuguiden som ble utarbeidet i forkant av opptaket for 2011: ”... å innhente og fortolke informasjon om søkeren...”. Dette skal videre danne grunnlaget for å sette de to prognosene: *Skoleprognose* og *Lederprognose*.

---

<sup>2</sup> Intervjuguide er utarbeidet i desember 2010 for opptaket i 2011. Elektronisk versjon mottatt elektronisk av Oblt Berg, Svein Olav ved FHS.

Skoleprognosen settes på bakgrunn av vurderinger av (se Figur 2-2):

- Motivasjon for valg av utdanning, hvor godt gjennomtenkt valget er, og hvor god innsikt vedkommende har i den utdannelsen han/hun har søkt
- Tidligere skolerresultat og selvrapporterte studievaner
- Alminnelig evnenivå, dersom denne informasjonen er tilgjengelig
- Fysisk form og antatt robusthet
- Annen relevant informasjon som fremkommer under intervjuet.



*Figur 2-2: Modell av skoleprognosen*

Skolepoeng fra videregående skole danner basis for skoleprognosen. Intervjuoffiser tar utgangspunkt i tallverdien for skolepoeng. AE, resultater fra fysisk test og vurderinger basert på intervjuet benyttes som grunnlag for om prognosen skal settes opp eller ned i forhold til skolepoeng. Intervjuoffiser forsøker å avdekke hvorvidt kandidaten har et større potensiale enn det skolepoengene fra videregående skulle tilsi, eller om vedkommende har jobbet hardt for å nå resultatet. I tillegg forsøkes avdekket studievaner og motivasjon som ligger bak resultatet. Det er ingen vektning eller matematisk vurdering som ligger til grunn for de vurderingene over, men en totalvurdering utført av den enkelte intervjuoffiser.

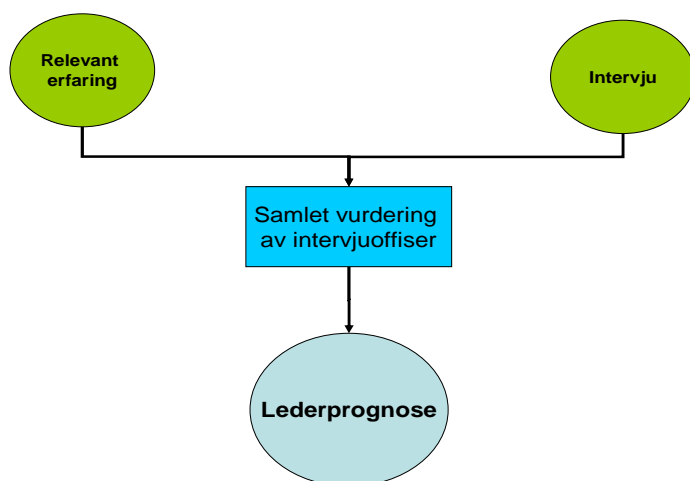
Samlet sett kan det være riktig å si at skoleprognosen settes på bakgrunn av tre objektive målinger og en mer subjektiv vurdering. AE angis på en ni-delt skala basert på et gjennomsnitt av flere ulike deltester. Fysisk testresultater angis i forhold til en fast skala som tar hensyn til kandidatens alder og kjønn, mens skolepoeng er beregnet på bakgrunn av samtlige karakterer og

fag fra videregående skole. Til slutt gjennomføres en samlet vurdering av de ulike elementene gjennom den enkeltes intervju av en erfaren intervjuoffiser. Her kan det variere noe fra intervjuoffiser til intervjuoffiser hvordan informasjon og fremtreden tolkes og vektlegges inn i den samlede prognosen.

Den andre prognosen som også settes på bakgrunn av intervjuet er lederprognosen.

Lederprognosen settes på bakgrunn av følgende momenter (se Figur 2-3):

- Relevant erfaring
- Refleksjonsnivå og modenhet
- Holdninger og verdier
- Stresstoleranse
- Kommunikasjonsnivå
- Selvinnsikt og personlighet
- Fremtreden under intervjuet



*Figur 2-3: Modell lederprognose*

De viktigste momentene under vurdering av lederprognose er evne til kommunikasjon og refleksjon. Her ser en etter hvordan kandidatene klarer å ordlegge seg både muntlig og skriftlig i forhold til å beskrive seg selv og sine tidligere erfaringer. Lederprognosen settes i løpet av samme intervju som skoleprognosen fastsettes. Som for skoleprognosen, er det her også fokus på potensiale, og ikke nåværende status.

Etter endt opptaksuke blir det gjennomført en foreløpig seleksjon på hvem som skal bli med videre inn i neste trinn av opptaket – feltøvelsen. Seleksjonen i denne fasen baseres på resultater fra fysiske tester, medisinske tester og intervjuet (med de to prognosene).

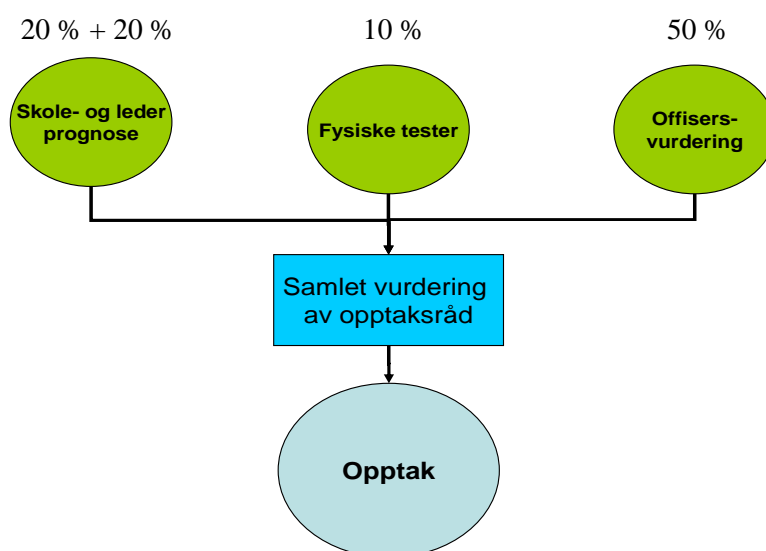
### 3.3 Feltøvelse

Neste fase er feltøvelsen. Kandidatene blir testet under feltforhold i ulike caser som er relevante for befalsyrket. Kandidatene settes sammen i lag som har med seg en veileder, og det rulleres på hvem av kandidatene som settes i leder-rollen i løpet av denne perioden.

Veiledere fører kontinuerlig logg på sine observasjoner og lager en samlet vurdering av den enkelte kandidat etter endt øvelse. Denne vurderingen benevnes offisersvurdering FOS. Det gis en skriftlig uttalelse på fem utviklingsområder. Disse er utvikling av lederrolle, utvikling av lederegenskaper, utvikling av selvforståelse, utvikling av mestringstro og utvikling av holdninger. Når det gjelder lederegenskaper som er nevnt som en av utviklingsområdene over er det følgende: evne til å ta initiativ, evne til å håndtere usikkerhet, evne til å skape tillit, omsorgsevne og evne til å ta selvstendige beslutninger (Forsvarets skolesenter, 2007, s. 17). Det gis en samlet vurdering i form av en hovedkarakter.

### 3.4 Hovedopptak

Det endelige opptaket til elev ved befalskolen skjer etter endt feltøvelse og er en vurdering basert på spesielt fire prediktorer: Skole- og lederprognose (vektes 20 + 20), fysisk form (vektes 10) og offisersvurdering FOS fra feltøvelsen (vektes 50). Dette er illustrert i Figur 2-4.



Figur 2-4: Modell opptak til elev etter gjennomført FOS BS

## 4 Metode

### 4.1 Deltagere

Det er kun ett befalsskolekull i Hæren (HBS kull 2010-2012) som både har blitt vurdert med en skole- og lederprognose ved FOS, og som har gjennomført skoleåret i etatsutdanningen med endelige skolerresultat. Fra et søkertall på 5195 personer til FOS BS Sommer 2010 ble det sendt innkalling til 685, hvorav 513 møtte til opptaket. Videre ble det tatt opp 229 elever etter endt opptak. Av de er det 194 elever som er uteksaminert etter endt skoleperiode. Fordeling av kjønn for denne gruppen er 88 % menn og 12 % kvinner. Gjennomsnittsalder for utvalget er på 19 år for menn og 20 år for kvinner. Denne undersøkelse vil ta utgangspunkt i alle kandidatene i utvalget som har fullført skoleåret med et oppnådd skolerresultat.

Avskalling fra antall søkere til antall innkalte er størst, og baserer seg på vurdering av skolepoeng fra videregående, AE og om de har plettfri vandel. Det er Vernepliktsverket som gjennomfører denne seleksjonen. Differansen mellom antall innkalte og oppmøte skyldes forhold hos den enkelte kandidat. Noen har fått andre tilbud som skoleplass på høyskole/universitet, eller at de har skiftet mening siden de søkte. Studien undersøker ikke bakgrunn for dette frafallet, men det vil være rimelig å anta at det både skyldes en del tilfeldige faktorer og antakelig manglende motivasjon. Frafallet fra oppmøte frem mot opptak som elev skjer på bakgrunn av seleksjonen som beskrives i denne studien. Når det gjelder frafall fra status som elev og frem mot fullført skoleår, kan det være flere årsaker. Noen velger å avslutte utdanningen selv. Det kan være at de finner ut at denne utdanningsretningen ikke var det de hadde trodd, eller manglende motivasjon. I tillegg skyldes noe av frafallet fysiske skader som oppstår i tjenesten, eller at tidligere skader blir fremprovosert av harde fysiske belastninger. I noen tilfeller kan frafallet skyldes manglende sikkerhetsklarering. Optimalt burde denne være klar før selve opptaket, men i praksis er det flere som ikke blir avklart før senere i utdanningsløpet. Frafall fra opptak til elev til endt utdanning er for dette kullet 15 %.

### 4.2 Prediktorer

De prediktorer som ble benyttet i denne oppgaven var skolepoeng fra videregående skole, AE, skole- og lederprognose fra intervjudelen, offisersvurdering FOS, samt total fysisk karakter fra FOS.

*Skolepoeng fra videregående skole* beregnes ut fra snittkarakter på vitnemål, samt tilleggspoeng for realfag eller språk. Det følger samme standard og beregninger som for opptak til annen høyere utdanning.



*Alminnelig evnenivå* er gjennomsnitt av flere evnetester benyttet i en årrekke av Forsvaret. Her gis det en skåre fra 1-9 basert på resultater fra tre prøver, numerisk -, generell - og verbal resonneringsevne.

*Skole- og lederprognosen* følger en skala fra 1-6, hvor 6 er beste skåre. Den fastsettes på bakgrunn av et semi-strukturert intervju, og skal gi en prognose for den enkelte kandidats potensiale innen lederskap og akademiske prestasjoner. Skoleprognosen tar utgangspunkt i skolepoeng fra videregående, AE, fysisk test og generelt inntrykk fra intervjuet. Lederprognosen baserer seg på eventuell tidligere erfaring og inntrykk fra intervjuet.

*Offisersvurdering FOS* er et eget vurderingsskjema hvor det gis en skriftlig vurdering innen de områder som er nevnt tidligere i studien. I tillegg gis det en samlet vurdering i form av en hovedkarakter med en skala fra A-F. I studien blir denne omregnet til en tallverdi med skala 0-5, hvor 5 tilsvarer beste karakter.

*Fysisk karakter fra FOS* består av en styrkedel med tre tester og en utholdenhetsdel med 3000 m løp. Det gis en samlet karakter beregnet ut fra gjennomsnitt av testene. Det benyttes en gradert skala fra 1 – 6, hvor 6 er beste skåre.

### 4.3 Kriterier

Utdanningen ved Hærens befalsskole foregår primært ute ved de respektive avdelinger i Hæren. Det gjennomføres noen samlinger eller moduler for elevene sentralt ved HBS på Rena i de tyngste teorifagene, herunder lederskap, instruksjonslære, militær idrett og trening, holdningsfag og militære fellesfag.

Selv om elevene får mesteparten av sin utdanning ute ved avdeling er det HBS sentralt som står for eksamensgjennomføring. Det sikrer lik håndtering både av vanskelighetsgrad og ikke minst sensur. Det avvikles fire teorieksamener ved skolen, samt praktisk eksamen i instruksjon og fysisk. I tillegg får hver enkelt elev en offisersvurdering. Det er en samlet vurdering av fremtreden og utvikling, samt utviklingspotensiale hos den enkelte elev.

Ved utstedelse av vitnemål ved HBS er de ulike delkarakterene vektet i henhold til en gitt standard. Kriteriene som ble nyttet i denne studien var hovedkarakter HBS, gjennomsnitt teorieksamener og offisersvurdering HBS.

Det er *hovedkarakter HBS* som bestemmer rangering av elever i kullet, og som avgjør hvem som blir beste elev. Den blir beregnet ut fra en vekting mellom de ulike fagene ved befalskolen. Den følger en gradert skala fra 0 – 5, hvor 5 er beste skåre.

*Gjennomsnittskarakter for alle teorifag* er ikke et kriterium som befalskolen benytter som noe selvstendig måleparameter. Den er beregnet eksplisitt for denne studien. Den ble fastsatt med

bakgrunn i gjennomsnitt for alle teoretiske eksamener som avlegges i utdanningsåret. Den følger en gradert skala fra A-F, hvor A er best. Denne variabelen er så konvertert til en gradert tallskala fra 0 – 5, hvor 5 tilsvarer beste skåre.

*Offisersvurdering HBS* baserer seg på det samme skjema som benyttes ved opptaket. Forskjellen ligger i at det nå etableres på bakgrunn av de prestasjoner som kandidatene har vist i løpet av utdanningsåret. Den følger samme skala som offisersvurdering FOS.

Årsaken til at studien benyttet snittkarakter teorifag og offisersvurdering HBS er at de representerer to faktorer som en spesifikt ser etter under seleksjonen, nemlig evne og potensial for å tilegne seg teori (akademisk prestasjon) og evnen til å utvikle seg som leder (lederpotensiale).

#### 4.4 Prosedyre

Resultater fra seleksjonen i form av testresultater og prognoser ble koblet sammen med senere prestasjoner innhentet i løpet av det første skoleåret. Opptaket ble gjennomført i juni med oppstart utdanning august samme år. Eksamensperioden ble gjennomført i mai påfølgende år. Datasettet med prediktorer er hentet fra personalsystemet, og Sjefspsykologen i Forsvaret har hatt tilgang til å hente ut de nødvendige data til oppgaven. Datasettet med kriterier etableres og forvaltes av Hærens befalsskole i form av en skoleprotokoll med alle relevante opplysninger om resultat og vurderinger. Begge datasettene ble håndtert av Sjefspsykologen, og analysene ble gjennomført på aidentifiserte data. Prosjektet ble vurdert av Norsk samfunnsvitenskapelig datatjeneste AS (NSD) som ikke meldepliktig eller konsesjonspliktig etter personopplysningsloven (Norsk samfunnsvitenskapelig datatjeneste AS, 16 mar 2012).

#### 4.5 Statistiske analyser

Samtlige analyser ble gjennomført ved hjelp av statistikkprogrammet SPSS (versjon 19) – *Statistical Package for the Social Sciences*. Denne studien undersøkte om det er sammenheng mellom de ulike prediktorer som benyttes ved seleksjonen til befalsskolen og de kriterier som det måles på ved endt skoleår. Korrelasjonsanalyser er en vanlig metode å benytte i slike sammenhenger. Med utgangspunkt i den type variabler som undersøkelsen legger til grunn ble Pearson  $r$  valgt som korrelasjonsmål (Ringdal, 2009, s. 299-300). Sammenhengen mellom prediktorene (testene og prognosene) og kriteriene (resultatet fra skoleåret) ble gjennomført ved hjelp av korrelasjonsanalyser og hierarkisk multippel regresjonsanalyse. Gjennom en korrelasjonsanalyse fremkommer styrken på sammenhengen mellom de ulike prediktorer og kriteriene, mens ved regresjonsanalyse belyses i hvor stor grad flere prediktorer til sammen

predikerer et kriterium og hvor stor del av variansen som forklares i den avhengige variabelen (Martinussen & Hunter, 2008, s. 118). På denne måten vil korrelasjonsanalysen bidra med å fortelle hvor sterk sammenheng det er mellom hver enkelt prediktor, og de kriteriene som er valgt. Slik sett ville en korrelasjonsanalyse alene kunne gitt gode svar på de spørsmål som stilles i problemstillingen. Allikevel vil en regresjonsanalyse tilføre en mer grundig analyse og forståelse av den samlede prediktive validitet til prediktorene. Det eksisterer flere metoder for å gjennomføre en regresjonsanalyse på (Johannessen, 2003; Pallant, 2010). I denne studien ble hierarkisk multippel regresjonsanalyse valgt som metode. Fordelen med denne type analyse er at en i tillegg til å få frem hvor mye av total varians som forklares av de ulike prediktorer, kan få estimert forklart varians på ulike trinn i prosessen. Slike vurderinger klarer man ikke å få frem ved en ren korrelasjonsanalyse eller en standard regresjonsanalyse.

Kandidater med manglende data (*missing*) var få og utgjorde ikke en stor prosentandel. Det er ingen grunn til å mistenke at det er noen systematikk i hvem som mangler data. Det skyldes mest sannsynlig noe mangelfulle rutiner for registrering. Det gjelder først og fremst for variablene skolepoeng videregående og AE. Analysene ble gjennomført ved å benytte ”*pairwise missing*”.

## 5 Resultater

### 5.1 Korrelasjonsanalyse

Tabell I viser en oversikt over deskriptiv statistikk og korrelasjoner mellom de ulike prediktorer som benyttes ved opptaket ved befalsskolen, og kriteriene etter endt utdanning i form av hovedkarakter HBS, snittkarakter teorifag og offisersvurdering HBS.

Tabell I. Gjennomsnitt, standardavvik og bivariate korrelasjoner for variablene i studien ( $N = 183 - 229$ )

Variabler	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8
<b>Prediktorer</b>										
1. Alminnelig Evnenivå	6.17	1.09								
2. Skolepoeng fra VGS	42.59	5.05	.32**							
3. Intervju Lederprognose	4.45	0.92	.08	.14*						
4. Intervju Skoleprognose	4.48	0.86	.27**	.53**	.60**					
5. Off vurd FOS	4.28	1.03	-.08	.05	.18**	.17**				
6. Fys tot FOS	3.31	0.75	.03	.12	-.02	.14*	.09			
<b>Kriterier</b>										
7. Hovedkarakter HBS	3.29	0.68	.26**	.35**	.22**	.38**	.23**	.10		
8. Snittkar Teorifag HBS	3.30	0.85	.23**	.47**	.17*	.41**	.05	-.01	.83**	
9. Offisersvurdering HBS	3.22	0.99	.15*	.00	.15*	.14*	.33**	.09	.65**	.16*

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$  (to-halet)

Alminnelig evnenivå og skolepoeng vgs. korrelerte henholdsvis  $r = .26$  og  $r = .35$  med kriteriet hovedkarakter HBS, mens skoleprognosen korrelerte med  $r = .38$ .

Videre korrelerte AE og skolepoeng vgs. med  $r = .23$  og  $r = .47$  med kriteriet snittkarakter teorifag HBS, og skoleprognosen korrelerte med  $r = .41$ .

Tabell I viser videre at lederprognosen korrelerte med kriteriet hovedkarakter HBS med  $r = .22$ , mens samme prognose korrelerte noe lavere med offisersvurdering HBS med  $r = .15$ .

Offisersvurdering FOS som er en samlet vurdering etter feltøvelsen (vurderingssentret) spesielt rettet mot lederskapspotensialet, korrelerte med  $r = .23$  opp mot kriteriet hovedkarakter HBS.

Samme prediktor korrelerte med kriteriet offisersvurdering HBS med  $r = .33$ .

Til slutt viser tabell I at den totale karakteren for fysisk test ved FOS var ikke signifikant korrelert med noen av kriteriene.

### 5.2 Multipl hierarkisk regresjonsanalyse

For å kunne vurdere i hvor stor grad de prediktorer som benyttes ved opptaket samlet sett predikerte skolerresultater ble det foretatt en hierarkisk multipl regresjonsanalyse. Variabelen Fysisk test FOS ble ekskludert fra videre analyser fordi korrelasjonsanalysen viste at prediktoren fremstod med ikke-signifikante korrelasjoner sett opp mot kriteriene.

I den første analysen ble prediksjon av skoleresultat/akademisk prestasjon analysert. I denne analysen ble skolepoeng vgs. og AE først vurdert, før skoleprognosen fra FOS ble tatt inn i steg 2.

Tabell II. Hierarkisk multipl regressjonsanalyse for akademisk prestasjon/skoleresultat.

Variabler	Akademisk prestasjon/skoleresultat	
	$\Delta R^2$	$\beta$
Steg 1	.23***	
Skolepoeng vgs		.34***
AE		.07
Steg 2	.03**	
Skoleprognose FOS		.22**
$R^2$	.26**	
$N$	192	

Note: Alle koeffisienter er tatt fra siste steg i regressjonsanalysen.

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

Tabell II viser at skolepoeng vgs. og AE forklarer 23 % av variansen for akademiske prestasjoner. Hele modellen med alle tre prediktorer inkludert, forklarer 26 % av variansen. Skoleprognosen forklarer 3 % i tillegg etter at det er kontrollert mot de to første prediktorer. Både skolepoeng vgs. og skoleprognose FOS var statistisk signifikante med henholdsvis  $\beta = .34$  og  $\beta = .22$ .

Videre ble prediksjon av lederpotensiale analysert. I regressjonsanalysen under ble de prediktorer som eksisterer før selve opptaket ved FOS satt inn i steg 1. I steg 2 ble de prediktorer som etableres under opptaket satt inn.

Tabell III. Hierarkisk multipl regressjonsanalyse for lederpotensiale inkludert grunnlagsdata før FOS.

Variabler	Lederpotensiale	
	$\Delta R^2$	$\beta$
Steg 1	.03	
Skolepoeng vgs		-.09
AE		.20**
Steg 2	.13***	
Lederprognose FOS		.08
Offisersvurdering FOS		.34***
$R^2$	.15***	
$N$	192	

Note: Alle koeffisienter er tatt fra siste steg i regressjonsanalysen.

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

Tabell III viser at de prediktorer som foreligger før selve opptaket starter ved Kjevik forklarer 3 % av variansen for offisersvurdering HBS eller lederpotensiale. Lederprognose FOS og offisersvurdering FOS forklarer i tillegg 13 % når de er kontrollert mot skolepoeng vgs. og AE.

Hele modellen forklarer 15 %. I denne analysen var det AE og offisersvurdering FOS som var statistisk signifikante med  $\beta = .20$  og  $\beta = .34$ .

I den siste regresjonsanalysen ble prediksjon opp mot hovedkarakter HBS vurdert. Denne karakteren er som tidligere nevnt en vektet skåre hvor de fag som ansees som viktigst teller mest i den totale sammenhengen. I steg 1 ble skolepoeng vgs. og AE lagt inn, mens i steg 2 ble de to prognosene fra intervjuet, samt offisersvurdering FOS lagt inn.

Tabell IV. Hierarkisk multippel regresjonsanalyse for total prestasjon inkludert grunnlagsdata før FOS.

Variabler	Total prestasjon/Hovedkarakter	
	$\Delta R^2$	$\beta$
Steg 1	.14***	
Skolepoeng vgs		.18*
AE		.16*
Steg 2	.08***	
Skoleprognose FOS		.19
Lederprognose FOS		.04
Offisersvurdering FOS		.19**
$R^2$	.23***	
$N$	192	

Note: Alle koeffisienter er tatt fra siste steg i regresjonsanalysen.

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

Tabell IV viser at skolepoeng vgs. og AE forklarer 14 % av den totale variansen for hovedkarakter HBS. Videre forklarer hele modellen 23%. De tre siste prediktorer (de som etableres under selve opptaket) forklarer 8 % i tillegg, når kontrollert for skolepoeng vgs. og AE. I den totale modellen var det skolepoeng vgs., AE og offisersvurdering FOS som var statistisk signifikante.

## 6 Diskusjon

Etter å ha presentert funnene i studien følger nå en diskusjon rundt de viktigste funnene. Først diskuteres korrelasjonsanalysen. Deretter følger en diskusjon rundt regresjonsanalysen med hovedvekt på å vurdere hvilke av de ulike prediktorer som samlet sett forklarer mest av variansen for de ulike kriteriene. Diskusjonen vil ta utgangspunkt i selve problemstillingen som går på sammenhengen mellom de ulike prediktorer og de valgte kriterier. Videre diskuteres hvor mye total varians forklares av de ulike prediktorer når de sammenstilles i en regresjonsanalyse. Funn i denne studien vil bli sammenlignet med tidligere undersøkelser innen forskningen. Avslutningsvis diskuteres styrker og svakheter ved denne studien.

### 6.1 Korrelasjoner

#### 6.1.1 Skolepoeng videregående skole og skoleprognose

Skoleprognosen som settes på bakgrunn av intervjuet korrelerer .41 med snittkarakter teorifag ved HBS. Videre korrelerer prognosen .38 med hovedkarakter ved HBS. Det harmonerer relativt bra med resultater fra tidligere meta-analyser utført der intervjuers prediktive validitet undersøkes. I en meta-analyse som blant annet ser på forskjellen mellom strukturerte og ustrukturerte intervju kommer de frem til en gjennomsnittlig korrigert validitet på .44 for strukturerte og .33 for ustrukturerte intervju med kriteriet jobbutførelse. Videre kom de frem til en validitet på .34 for strukturerte og .36 for ustrukturerte når kriteriet var evnen til å tilegne seg ny kunnskap og ferdigheter (McDaniel et al., 1994, s. 606-608). Intervjuene under opptaket på FOS gjennomføres som delvis strukturerte intervju, og resultatene harmonerer således relativt godt med tidligere funn. Nå er det imidlertid flere forskjeller mellom denne studien og den som det refereres til. For det første tar den for seg et tradisjonelt jobbintervju, mens på opptaket til befallsskolen ser en først og fremst etter potensiale for akademiske prestasjoner. Når det gjelder valg av kriterium er det også en forskjell ved at i denne studien benyttes karakterer fra eksamener og vurderinger, mens det i den refererte studien benyttes vurderinger av jobbutførelse og opptreningsevne. I en annen studie kommer strukturerte intervju ut med en prediktiv validitet på .51, mens ustrukturerte har .38, når evalueres opp mot total jobbprestasjoner (Schmidt & Hunter, 1998, s. 265). Det kan synes å være noe høyere enn hva som fremkommer i denne studien. Forskjellen kan derimot sies å være relativt lik dersom en analyserer det grundigere. For det første er intervjuet på FOS et delvis strukturert intervju. En kan da forvente at resultatet bør ligge mellom .38 og .51, noe det også gjør opp mot akademisk prestasjon. For det andre er skoleprognosen ment å skulle forutsi akademiske prestasjoner, og her viser den til en korrelasjon på .41. At den samme prognosen korrelerer svakere mot total prestasjon (.38), følger naturlig av

at det kriteriet i tillegg inneholder vurderinger som prognosen ikke er ment å predikere. En annen forskjell mellom denne studien og meta-analysene er at de ofte presenterer korrigerte korrelasjoner, dvs. der de gjennomsnittlige korrelasjonene er korrigert for manglende reliabilitet i kriteriet og *restriction of range* – problemet noe som gjør dem høyere enn ukorrigerte korrelasjoner som presenteres i denne studien.

Skoleprognosen tar som tidligere nevnt utgangspunkt i skolepoengene fra videregående skole. I tabell I korrelerer de to innbyrdes med .53, noe som kan tyde på at skoleprognosen bygger på skolepoeng fra videregående. Det beskrives også i prosedyren eller intervjuguiden i forbindelse med FOS BS. Undersøkelsen understøtter derfor at intervjuoffiserene gjennomfører dette i praksis. Likevel viser den samme korrelasjonen at skoleprognosen er basert på andre vurderinger og således ikke bare er en ren konvertering fra skolepoeng til prognose. Det som imidlertid er litt interessant i analysen er at skolepoeng fra videregående skole korrelerer med snittkarakter teorifag ved HBS med .47, altså en høyere korrelasjon enn det som oppnås etter intervjuet. Det er nærliggende å kunne forvente at skoleprognosen skulle ha hatt en større korrelasjon enn grunnlaget den baserer seg på. Det kan være flere årsaker til det. Det kan være slik at den enkelte intervjuoffiser ikke evner på en god nok måte å tolke og vurdere den informasjonen som fremkommer under intervjuet. I hvert fall i forhold til å kunne forutsi hvordan den enkelte kandidat vil kunne prestere akademisk. En annen forklaring kan være at de spørsmål som stilles og den informasjonen som fremkommer, ikke er formulert på en slik måte at de egner seg til å styrke prediksjon av akademiske evner. Imidlertid korrelerer skolepoeng videregående .35 med hovedkarakter HBS som er tilnærmet likt som for skoleprognosen (.38). Begge disse prediktorer korrelerer svakt med offisersvurdering HBS med .14 (skoleprognose) og .00 (skolepoeng). Her må det da tilføyes at ingen av de to prediktorer er i utgangspunktet ment å skulle predikere potensiale for lederskap. Samlet sett kan det dermed se ut for at både skolepoeng vgs. og skoleprognosen FOS har en relativt høy korrelasjon med både akademisk evne og total prestasjon ved HBS. Funnene i denne studien hva gjelder skolepoeng fra videregående sammenfaller med andre studier innen seleksjon. I en studie som ser på rangering fra videregående (*ranking High School*) fra *US Navy* (Alf et al., 1988, s. 6) viser de til en korrelasjon mellom ”skolepoeng” og akademiske evner med .44. Videre korrelerer samme prediktor med total prestasjon med .35. Med andre ord nesten helt identiske funn sett opp mot korrelasjonene i denne studien. Rammene rundt studien fra *US Navy*, samt prediktor og kriterier er da også veldig like. Seleksjon inn til Forsvaret, resultat fra videregående og kriterier på henholdsvis akademiske resultat og total prestasjon er klare likhetstrekk.



Når det gjelder reliabilitet for skoleresultat som prediktor er det rimelig å anta at den har høyere reliabilitet enn for eksempel vurderinger av observatør eller selv-vurderinger (Alf et al., 1988; Jones, 1989). Imidlertid kan det være grunn til å være kritisk til reliabiliteten til skole- og eksamensresultater også. Standardisering, prosedyrer og skåringsregler kan nok sies å være godt utviklet for en slik variabel. Det forutsetter imidlertid et vel etablert regime rundt gjennomføring og sensur i form av sensorrettledninger, og at det er flere sensorer per oppgave. På en annen side kan for eksempel motivasjon og dagsform hos den enkelte kandidat gi utslag. Motivasjon kan gi utslag både i forhold til eksamenskarakterer og standpunktkarakterer. Det er ikke alle som i en alder av 17-20 år har et reflektert forhold til fremtidig utdanning og karriere. De kan da underprestere i forhold til egentlig potensiale, og dermed ha store utviklingsmuligheter i et for dem stimulerende miljø. Når det gjelder eksamensresultat kan en faktor som dagsform og tema for eksaminasjon gjøre utslag som ikke nødvendigvis viser "riktig" karakter. For å fange opp slike forhold kan intervjuet være en god arena. Her har intervjuoffiser mulighet til å avdekke om det ligger slike forhold bak de resultater som fremlegges ved opptaket. Slik det er beskrevet i intervjuguiden for intervjuoffiser er det nettopp slike forhold de blant annet skal forsøke å avdekke. En kandidat som har et høyt snitt fra videregående kan ha oppnådd det på forskjellige måter. Noen har jobbet strukturert og hardt for å oppnå denne karakteren, mens andre kan ha oppnådd det samme resultatet på en enklere måte. Slike forhold kan si noe om potensiale til kandidaten. På den måten bør intervjuet kunne styrke prediksjonen av blant annet akademiske evner. Ut fra korrelasjonsanalysen i denne studien kan det se ut som prognosen fra intervjuet ikke i stor grad klarer å styrke prediksjonen av akademiske prestasjoner. Den kommer ut med en litt svakere korrelasjon enn skolepoeng fra videregående. Forskjellen mellom korrelasjonene er ikke signifikant. Oppsummert kan det tyde på at både skolepoeng fra videregående og skoleprognosen fra intervjuet kan sies å være en god prediktor for fremtidige akademiske prestasjoner. De viser i tillegg en god prediktiv validitet for total prestasjon. Det er naturlig at den siste blir noe lavere, da den inneholder andre parametre i tillegg ut over akademiske evner. Det kan være viktig å ta med seg videre at de to prediktorer er først og fremst egnet til å forutsi akademiske evner, og ikke en generisk prediktor til befalsskolen.

### 6.1.2 Alminnelig evnenivå

AE korrelerer litt svakere enn de foregående prediktorer med .23 med snittkarakter teorifag HBS, og .26 med hovedkarakter HBS. Videre korrelerer AE enda svakere med offisersvurdering HBS (.15). I en studie om seleksjon av offiserer til det sveitsiske forsvaret vises det til en korrelasjon for AE på .26 med akademisk evne og .21 med militær prestasjon (Melchers & Annen, 2010, s.

110). Spesielt for akademisk evne sammenfaller funn i denne oppgaven med nevnte studie. Den største differansen er opp mot militær prestasjon. Det kan nok skyldes blant annet at dette kriteriet ikke er direkte sammenlignbart i de to studiene. Offisersvurdering HBS er vinklet spesifikt mot lederskapspotensiale, mens den sveitsiske studiens kriterie legger til grunn en mer helhetlig og bredere prestasjon innen militære fag og disipliner. Videre vises det til relativt høy validitet for AE både i forhold til å predikere fremtidig jobbutførelse (.23-.27) og fremtidig opptreningsevner (.23-.29) i en meta-analyse fra 2003 (Salgado et al., s. 1075). Spennet i validitet som fremgår i parenteser viser til forskjeller i kompleksitet som det differensieres på i analysen. En studie som imidlertid ser på AE som en prediktor for lederskap spesielt, viser til en korrelasjon på .21 (Judge, Colbert, & Ilies, 2004). Alt sett under ett kunne en derfor ha forventet en noe større korrelasjon mellom AE og offisersvurdering HBS ved opptaket til befalsskolen. Nå kan det tenkes at korrelasjonen hadde blitt høyere dersom en hadde valgt et kriterium noe senere i karrieren, og ikke etter kun ett år ved befalsskolen. Selv om testene som benyttes for å avdekke AE er ansett som rimelig valide og reliable tester, kan det her også være ulike feilkilder. Testen gjennomføres på sesjon tidligere i prosessen. Ikke alle ungdommer som møter her har et bevisst forhold til Forsvaret eller konkrete planer om å søke befalsskolen. Motivasjonen for å prestere under en slik test kan derfor variere i stor grad. I tillegg kan forhold som dagsform gi utslag. Slike forhold kan være vanskelig å kontrollere i forbindelse med en slik test. Det gis heller ikke i utgangspunktet adgang til å gjennomføre en re-test. Kun helt spesielle forhold kan medføre at dette innvilges. Allikevel er det rimelig å kunne anta at de fleste presterer i forhold til potensiale, og at situasjoner som nevnt over er tilfeldig og i sterkt mindretall.

### 6.1.3 Lederprognose

Når det gjelder lederprognosen er den ment å predikere lederpotensialet hos kandidatene. Offisersvurdering HBS er en vurdering av utøvelse av lederskap og lederpotensiale etter ett års skolegang. Imidlertid er ledervurdering bare en del av flere elementer som inngår i kriteriet hovedkarakter HBS. Av analysen fremkommer det at lederprognosen korrelerer .15 med offisersvurdering HBS, mens den samme prediktoren korrelerer .22 med hovedkarakter HBS. Lederprognosen korrelerer faktisk (marginalt) litt høyere opp mot snittkarakter teorifag (.17) enn den gjør mot offisersvurdering HBS. Tar en utgangspunkt i hensikten med prognosen, kan det tyde på at den fremstår som noe svak i forhold til det den er ment å skulle predikere. På en annen side er det kanskje ikke så merkelig, da den settes under samme intervju som skoleprognosen settes. Disse to prognosene korrelerer innbyrdes med .60, og det kan dermed tyde på at lederprognosen kan være noe påvirket av skoleprognosen. Skoleprognosen tar utgangspunkt i en

allerede gitt karakter, nemlig skolepoeng og AE, mens lederprognosen må dannes på bakgrunn av den korte samtalen som finner sted under intervjuet. Sammenlignet med tilsvarende undersøkelser som nevnt over i forbindelse med intervju, skulle en kunne forvente en noe høyere korrelasjon mellom prediktoren lederpotensialet og spesielt offisersvurdering HBS. Det er imidlertid en krevende oppgave å vurdere og å gi en presis vurdering av potencialet i løpet av et kort intervju. Tidligere studier viser til flere aspekter som kan være med å påvirke resultatet. Blant annet diskuteres ofte selve intervjueren i forhold til halo-effekt og evnen til å unngå forutinntatthet. Med halo-effekt menes at intervjuoffiserer vurderer kandidatene som generelt gode eller dårlige, og at denne generelle følelsen er med å prege de videre vurderinger av kvaliteter (Nathan & Tippins, 1990, s. 290). Videre diskuteres rammen for selve intervjuet i form av omgivelser, dagsform og ikke minst tidsaspektet (Zickar, Cortina, & Carter, 2010, s. 400-401). Samtidig viser annen forskning på at halo-effekten kan være noe overdrevet. I studien til Nathan og Tippins fremheves det at denne effekten ikke er en feilkilde, men derimot kan være en styrke i forbindelse med rangering av kandidater (Nathan & Tippins, 1990). I studien viser de til at intervjuere som var meget bevisst forholdet rundt halo-effekt kunne ha en tendens til å vektlegge små negative hendelser, som igjen gjorde utslag i deres vurdering. De intervjuere som hadde vurderinger som var farget av halo-effekt hadde en tendens til å kunne overse slike små hendelser og klarte dermed å fokusere på en total vurdering av kandidaten. Undersøkelsen understøttes av en tidligere studie om dette fenomenet (Kahneman & Tversky, 1973). Selv om det gjennomføres utvelgelse av intervjuoffiserer, gis ut intervjuguide og gjennomføres kompetanseheving blir de to prognosene en subjektiv vurdering av hver enkelt kandidat. Det er flere forhold som spiller inn med hensyn til fastsetting av den enkelte prognose. For det første vil hver enkelt intervjuoffiser ha ulik bakgrunn og derfor analysere ulikt med hensyn til å vurdere og å fastsette en prognose. For det andre kan dagsform og opplevd miljø påvirke hver enkelt kandidats evne til å kommunisere i løpet av et slikt kort intervju. Et tredje aspekt er hvorvidt intervjuoffiseren klarer å skape en god flyt og trygghet i en slik, for de fleste, uvant situasjon.

Ved vurdering av reliabilitet i forbindelse med intervju skilles det ofte mellom fire komponenter. Det er kongruens, presisjon, objektivitet og konstant (Trost & Jeremiassen, 2010, s. 137-141). Kongruens handler om likhet i spørsmålsstilling og det å måle det samme. Presisjon gjelder evnen til å registrere svar. Objektivitet dreier seg om intervjuerens evne til å registrere like forhold på samme måte. Og til sist konstant, som henger sammen med at holdninger og det som fremkommer underveis i intervjuet ikke endrer seg over tid. I og med at intervjuet under seleksjonen på FOS BS er å anse som et semi-strukturert intervju kan det stilles spørsmål ved

kongruens. Det er åpning for å bygge på eller trekke fra spørsmål alt etter gangen i intervjuet og intervjuoffiserens tanker. Allikevel er det lagt opp til at hvert intervju i grovt skal innom de forslag til spørsmål som ligger i intervjuguiden. Presisjon avhenger i stor grad av evnen til intervjuoffiser til å notere momenter og svar underveis. Det nyttes ikke opptaker eller video av intervjuene. Det er derfor helt avgjørende at intervjuoffiser er våken og har evnen til å registrere alt fra muntlige svar og eventuelt kroppsspråk av betydning. Samtidig med at en skal få et slikt intervju til å flyte kan det være en krevende øvelse. Erfaring som intervjuoffiser kan være en viktig dimensjon rundt dette spørsmålet. Objektivitet kan være en utfordring ved intervju, og her kommer blant annet halo-effekt og forutinntatthet inn. Disse momentene er behandlet tidligere i studien. Når det gjelder forholdet som benevnes konstant, kan det være en utfordring spesielt ved å sette en lederprognose. Holdninger og forhold som avdekkes under et slikt intervju går ofte på personlighet og personlige egenskaper der og da. Med hensyn til at lederskap kan sies å være et modningsfag, og hvor de fleste kandidatene i sin helhet mangler erfaring innen denne retningen, kan forholdet rundt konstant være en utfordring. På en annen side er det potensialet som intervjuoffiseren er ute etter, og det man ser etter hos kandidatene kan sies å være grunnholdninger som danner grunnlaget for å kunne bli en dyktig leder.

Alt sett under ett kan det derfor sies å være flere faktorer som spiller inn hvorvidt en slik prognose kan anses som reliabel, noe som igjen vil påvirke den prediktive validiteten (Schmidt & Hunter, 1998). I denne studien viser skoleprognosen en relativt sterk korrelasjon med kriteriet den er ment å predikere, mens lederprognosen oppnår en svakere korrelasjon.

#### **6.1.4 Offisersvurdering FOS**

Når det gjelder offisersvurdering FOS synes den å korrelere relativt godt med både offisersvurdering HBS (.33) og hovedkarakter HBS (.23). Tidligere analyser understøtter at vurderingssenter fungerer relativt godt når det gjelder seleksjon. I studien fra det sveitsiske forsvaret vises det til en korrelasjon for vurderingssenter på .31 opp mot akademisk evne og .32 opp mot militær prestasjon (Melchers & Annen, 2010). Prediksjon av militær prestasjon er relativt lik, mens akademisk evne har en noe lavere prediktiv validitet i denne studien. Det kan skyldes flere faktorer, men en av de mest nærliggende antas å være det fokus som ligger til grunn for offisersvurdering FOS. Den er først og fremst rettet mot lederpotensiale, og tar i utgangspunktet ikke med vurderinger rundt akademiske evner for kandidatene. Videre viser Schmidt og Hunter i sin meta-analyse (1998, s. 265) til en validitet for vurderingssenter på .37. Imidlertid viser flere andre studier til en diskusjon rundt spesielt innholdsvaliditet for vurderingssenter (Lance, Woehr, & Meade, 2005; Woehr & Arthur, 2003). Under opptaket til

befalsskolen består feltøvelsen av en rekke ulike caser som skal løses av de enkelte gruppene. Planlegging og utvelgelse av de "riktige" casene kan være avgjørende for om vi får målt det vi ønsker. Casene evalueres og utvikles fra år til år. I tillegg foregår feltøvelsen over lang tid og inneholder et relativt bredt spekter av oppgaver. På denne måten er det rimelig å anta at oppgavene dekker det spekter av situasjoner som er nødvendig for å kunne gjøre gode observasjoner.

Offisersvurdering FOS som skrives på bakgrunn av prestasjoner under feltøvelsen er en samlet vurdering basert på flere forhold. Disse er nevnt tidligere i oppgaven. Når det gjelder vurdering av reliabilitet er det flere forhold som bør trekkes frem. Å skulle få samme resultat av de ulike vurderingene ved en ny måling kan oppleves som noe utfordrende. Miljøet som skapes gjennom denne perioden påvirkes av såpass mange forhold at det kan vanskelig karakteriseres som statisk. Gruppesammensetning, dagsform, spenninger i gruppen, vær og føreforhold for å nevne noen. På en annen side går denne perioden over såpass lang tid at personlige svingninger hos den enkelte kandidat bør kunne fanges opp, slik at "ekstremverdier" ikke blir dominerende for den endelige vurderingen.

Et annet forhold innen reliabilitet er hvor konform veiledning og vurdering gjennomføres. Det benyttes i utgangspunktet rutinerne veiledere på FOS BS. Rutinerne i form av godt kjent med befalsskolen, veiledning og vurderinger som daglig gjennomsyrrer denne type tjeneste. I løpet av opptaksuken gjennomfører veiledere og kompaniene<sup>3</sup> forberedelser for feltøvelsen. I tillegg til et ledelsesapparat inneholder denne strukturen nivellører. Det er spesielt utvalgte offiserer som har lang erfaring innen vurdering og veiledning. Funksjonen til de er å støtte hver enkelt veileder etter behov. Det gjennomføres daglige veiledningsmøter hvor erfaringer og observasjoner deles og diskuteres. I tillegg er det tilgang til psykologer fra Sjefpsykologen som er med ut på øvelsen og kan støtte ved behov. På denne måten forsøker en å sikre en lik og mest mulig rettferdig behandling og vurdering av kandidatene.

En ulempe med vurderingssenter er at de ofte strekker seg over lenger periode og er således en ressurskrevende målemetode. Ved FOS BS går denne perioden over én uke, med tillegg i lønn for det befalet som følger med ut. Investeringen i denne delen av opptaket kan sies å være relativt stor. Jones (1989, s. 72) antyder i sin studie at prosedyrer i forbindelse med vurderingssentre kan være lenger enn nødvendig, og at veilederne ikke klarer å nyttiggjøre seg all tilgjengelig informasjon. Imidlertid viser Jones i samme artikkel (1989, s. 73) til kost-nytte

---

<sup>3</sup> Kompanistruktur – HBS organiserer seg i kompani- og troppsstruktur med ledelsesapparat. Normalt ca 200-250 i hvert kompani, fordelt på 4-5 tropper i hvert kompani.

vurderinger som er foretatt, at bruk av vurderingssenter har en relativt høy nytteverdi i form av mer presis seleksjon. Noe av utfordringen med feltøvelsen blir derfor å finne en optimal balanse med tanke på lengde og innhold, både i forhold til kostnader og kvalitet på vurderingene. Alt sett under ett viser meta-analyser av vurderingssenter til relativt gode resultater når det gjelder validitet generelt (Petrides, Weinstein, Chou, Furnham, & Swami, 2010, s. 227). I tillegg viser tidligere meta-analyser som omhandler prediktiv validitet spesielt for vurderingssenter til relativt høy validitet (Hermelin et al., 2007).

#### 6.1.5 Fysisk test FOS

Fysisk test FOS er som beskrevet under opptaket en av de prediktorer som benyttes ved selve hovedopptaket til elev med en vektning på 10 %. Som det fremgår av tabell I viser denne prediktoren lav korrelasjon med kriteriene i studien. Det gjelder både i forhold til akademisk prestasjon, lederpotensiale og hovedkarakter. Funnet i denne studien sammenfaller med annen tilsvarende forskning som ser på sammenheng mellom fysisk test og senere jobbutførelse. I en undersøkelse for *US Navy* (amerikanske Marinen) viser de til liten sammenheng mellom fysiske tester og senere fysisk krevende jobbutførelse (Marcinik et al., 1995). Fysisk test i studien fra *US Navy* er nesten identisk med den som benyttes ved FOS BS. Likevel bør en være forsiktig med å trekke den konklusjonen at fysisk test ikke er av betydning for senere prestasjoner eller utøvelse av offisersyrket. Derimot kan det tyde på at resultater fra fysisk test FOS ikke har en direkte sammenheng med prestasjonsmålene som benyttes etter endt skoleår. En av årsakene kan være at fysisk test ved endt skoleår har en relativt lav vektning når det gjelder kriteriet hovedkarakter. Allikevel kan det være interessant å merke seg at det er meget liten sammenheng med akademiske evner og lederprestasjoner. I en studie fra Norges Idrettshøgskole antydes det at fysiske egenskaper ser ut til være overdrevent i fokus under spesielt feltøvelsen (Rones & Fasting, 2011). Imidlertid viser tabell I en korrelasjon på .09 mellom fysisk test FOS og offisersvurdering FOS. På den ene siden kan det tyde på at fysisk form ikke påvirker i så stor grad de samlede vurderinger som gjøres under feltøvelsen. På en annen side kan det også være at det er andre dimensjoner innen fysisk form som vektlegges under feltøvelsen enn de som testes under opptaksuken. I første rekke kan dette være egenskaper som robusthet og styrke i felt. Allikevel viser tidligere forskning på fysisk test ved seleksjon til at det finnes fordeler og utbytte av å benytte slike tester i seleksjon til fysisk krevende jobber. Den mest fremtredende nytten ser ut til å være mindre fysiske skader og økt produktivitet. Fysisk god form ser ut til å være et grunnlag for å unngå blant annet belastningsskader (Gebhardt & Baker, 2010, s. 294). Skoleåret ved befalsskolen er en hektisk periode, hvor både teori og fysisk krevende tjeneste er

fremtredende. For noen kan det være en bratt progresjon hva gjelder fysisk belastning med mye fysisk krevende aktivitet, kombinert med tung utrustning. Sett i lyset av denne problemstillingen vil det være riktig å kunne vise til viktigheten av en viss fysisk standard på kandidatene.

Allikevel viser resultatet fra fysisk test ved FOS tilnærmet ingen sammenheng med de kriteriene som benyttes i denne studien. Alt i alt kan det da være grunn til å stille spørsmålsteget hvorvidt fysisk test fra FOS fortjener å ha rollen som en selvstendig prediktor i det endelige opptaket med en vekting på 10 %.

#### **6.1.6 Reliabilitet og validitet for kriterier**

Etter å ha diskutert reliabilitet og validitet for de prediktorer som benyttes ved opptaket følger nå en vurdering av de samme momenter for de kriterier som benyttes i oppgaven. Schmidt et al. (1992, s. 656) viser til at diskusjonen som har pågått rundt valg av kriterier kanskje har vært noe overdrevet i forhold til vurderinger om spesielt kognitive testers validitet. Graden av korrelasjon vil kunne variere mellom de ulike kognitive tester og kriterier, men uavhengig av hvilke kriterier som velges har ingen kommet ut med ingen validitet. I denne studien vil dette gjelde i forbindelse med vurdering av validitet mellom AE og ulike kriterier.

Tradisjonelt vil det være riktig å antyde at skolekarakterer er å anse som mer reliabel enn vurderinger foretatt av for eksempel overordnede (Schmidt et al., 1992, s. 656). I denne oppgaven vil det si at en kan anta høyere reliabilitet for kriteriet snittkarakter teorifag ved HBS enn for kriteriet offisersvurdering HBS. Vurderingene ved utfylling av offisersvurdering HBS gjennomføres av overordnede som følger kandidatene i den daglige tjenesten. Feilkilder i form av forutinntatthet og fordommer kan være med å svekke reliabiliteten til dette kriteriet.

Konsekvensen av dårlig reliabilitet for et kriterium vil være en svakere prediktiv validitet enn om reliabiliteten hadde vært god. Vurderingen som foretas av de som følger kandidatene i den daglige tjenesten kunne blitt annerledes med andre overordnede. Flere studier peker på at vurderinger av jobbutførelse som et kriterium er problematisk (Murphy, 2008, s. 148). I tillegg viser samme studie til at det kan være vanskelig å avdekke når og hvor vurderingsfeil oppstår. Som nevnt tidligere viser på en annen side andre studier til at feilkilder som halo-effekt, forutinntatthet, vennskap og likheter ikke svekker slike vurderinger (Schmidt et al., 1992, s. 656-657).

En annen vurdering som bør diskuteres er om kriteriene måler det som er riktig og viktig i forbindelse med denne utdanningen. I denne sammenhengen er det nærliggende å kunne antyde at det er offisersvurdering som kan sies å fremstå som mest reliabel. Skjema og vurderinger er standardiserte og har vært utprøvd over lengre tid. Offisersvurdering HBS er heller ikke så utsatt

for dagsform og personlige svingninger, da den baserer seg på resultater prestert over en lengre tidsperiode. Teoretiske eksamener endrer oppgavesett hvert år, og kan således være vanskeligere å sjekke opp mot pålitelighet og nøyaktighet. Den vil i tillegg være sårbar i forhold til dagsform hos den enkelte kandidat. Allikevel vil det være riktig å kunne antyde at skolerresultat eller eksamenskarakterer normalt er å anse som et kriterium med høy reliabilitet.

## 6.2 Regresjonsanalyse

### 6.2.1 Akademisk prestasjon

Når det gjelder prediksjon av akademisk prestasjon, viser tabell II at mye av forklart varians ligger i de data som er innhentet før selve opptaket starter. AE og skolepoeng fra videregående skole gir et solid og godt grunnlag for å kunne forutsi kandidatens potensiale innen akademisk prestasjon eller kognitive evner. Allikevel tilfører intervjuet i form av skoleprognosen ytterligere presisjon i denne prediksjonen, selv om det må kunne sies å være noe begrenset. Som det er beskrevet tidligere bygger skoleprognosen på skolepoengene fra videregående. En fare ved en slik fremgangsmåte er at de to variablene blir så lik at de blir å anses som en dublering. Dersom så er tilfelle vil det være problematisk å inkludere begge i en regresjonsanalyse. Dette kalles multikolaritet. En indikasjon på det vil være en svært høy  $r$  mellom de uavhengige variablene. Korrelasjonen ligger i dette tilfelle på .53 mellom skolepoeng videregående og skoleprognosen. Ved høyere korrelasjon mellom to uavhengige variabler enn .70, er dette problematisk og et tegn på at de måler mye av det samme (Johannessen, 2003, s. 135). Så i dette tilfellet vil det være riktig å kunne anse de to variablene som to selvstendig bidragsyttere og ikke en vurdering av det samme.

Både skolepoeng videregående og skoleprognose fremstår som statistisk signifikante prediktorer. Selv om skoleprognosen ikke tilfører mye forklaring av total varians, fordi den er høyt korrelert med skolepoeng fra videregående, forklarer den allikevel ytterligere en signifikant andel av variansen for kriteriet. Nå skal det tillegges at dersom en hadde satt inn skoleprognosen inn i steg 1 hadde resultatet blitt snudd på hodet. Da hadde skoleprognosen vært den prediktoren som hadde forklart mest av total varians. Imidlertid er det slik at skolepoeng vgs. er den prediktoren som eksisterer allerede før opptaket starter, og det vil være naturlig at denne settes inn først i regresjonsanalysen.

Når det gjelder  $\beta$ -verdi (eller regresjonskoeffisient) som er statistisk signifikant, kan den benyttes til å predikere fremtidig skolerresultat (Ringdal, 2009, s. 363-364). For skolepoeng vgs. vil det si at økning på en verdi for den prediktoren, sannsynlig resulterer i at skolerresultatet øker med .34 ( $p < .001$ ). Det forutsetter at de andre variablene holdes konstant.



### 6.2.2 Lederpotensiale

Når det gjelder prediksjon av lederpotensiale ble de prediktorer som eksisterer før opptaket lagt inn i steg 1. Videre ble de prediktorer som etableres under selve opptaket lagt inn i steg 2. Tabell III viser at grunnlaget som eksisterer før opptaket forklarer relativt lite av total varians for lederpotensiale. Det er først og fremst lederprognosen fra intervjuet og offisersvurdering FOS fra feltøvelsen som forklarer mest av den totale variansen for lederpotensiale. Allikevel fremstår AE som en signifikant bidragsyter i forklaring av varians for lederskap. Skolepoeng fra videregående fremstår ikke som en sterk bidragsyter i denne modellen. Det kan vel ikke sies å være direkte overraskende. Skolepoeng fra videregående er i utgangspunktet ikke tenkt å være en prediktor for fremtidig lederskap ved opptaket. Skoleprognosen fremstår da også som en bedre prediktor for akademiske prestasjoner enn for militært lederskap i følge tidligere studier (Jones, 1989, s. 67). At AE fremstår som en unik bidragsyter i prediksjon av lederskap er sammenfallende med tidligere studier (Judge et al., 2004). Allikevel viser andre studier at AE predikerer bedre fremtidig jobbutførelse og evnen til å lære enn hva gjelder lederskap (Withman, 2009, s. 18). På denne måten understøttes funn i denne studien av tidligere forskning på området. På en annen side ser AE ut til å være en bedre prediktor av fremtidig lederskap enn hva skolepoeng fra videregående er. Når det gjelder de prediktorer som etableres under selve opptaket fremstår offisersvurdering som en unik bidragsyter i forklaring av lederskap. Lederprognosen fremstår som en svak prediktor. Det sammenfaller også med korrelasjonsanalysen, hvor lederprognosen hadde relativ lav korrelasjon med offisersvurdering HBS. Offisersvurdering FOS derimot ser ut til å styrkes som prediktor ved regresjonsanalysen. Tilføring av forklart varians for lederpotensiale, kombinert med høy  $\beta$ -verdi og statistisk signifikans viser at den fremstår som en prediktor av fremtidig lederskap med relativt høy prediktiv validitet. I denne studien forutsetter det derimot at offisersvurdering HBS er et godt kriterium for fremtidig lederskap. Det er kanskje det best tilgjengelige kriteriet for lederskapsprestasjoner etter ett år ved skolen. På en annen side kan lederskap sies å være et modningsfag, og et lederskapskriterium senere i karrieren hadde kunne gitt et bedre resultat. For det første kunne en måling senere i karrieren gitt et bedre kriterium, men det kunne i tillegg ha gitt et annet resultat hva gjelder prediktiv validitet for lederprognosen. Total forklart varians i Tabell III er på 15 %. Det er noe lavere enn tilsvarende studier viser. I en meta-analyse av vurderingssenter som ser på ulike dimensjoner internt i et vurderingssenter forklares 20 % av total varians for jobbutførelse (Arthur, Day, McNelly, & Edens, 2003, s. 143). Det er imidlertid noen vesentlige forskjeller på den studien og denne. Ved opptaket til befalsskolen er både prediktor og kriterium først og fremst rettet mot lederskap. I studien fra 2003 fokuseres det på generell jobbutførelse. I tillegg er det i Tabell III med flere

prediktorer i tillegg til vurderingssenter. Likheten derimot er at kriteriene baseres på vurderinger foretatt av blant annet overordnede. Således kunne en ha forventet en større forklart varians i denne studien.

### 6.2.3 Total prestasjon

Videre ble det gjennomført en regresjonsanalyse opp mot total prestasjon, eller hovedkarakter HBS i dette tilfellet. Prediktorer som eksisterer før opptaket ble lagt inn i modellen i steg 1, før prediktorer som etableres under opptaket ble lagt inn i steg 2. Samlet sett forklarer modellen 23 % av variansen for total prestasjon. Videre ligger det et godt grunnlag allerede i de prediktorer som befalsskolen har ved oppstart av opptaket. Skolepoeng vgs. og AE forklarer 14 % av variansen. Imidlertid tilføres det ytterligere forklaring av varians med de prediktorer som etableres under selve opptaket. I forhold til total prestasjon er det spesielt offisersvurdering FOS, skolepoeng videregående og AE som var statistisk signifikante og dermed kan sies å være unike bidragsytere. Som Tabell IV viser mister både skoleprognosen og lederprognosen noe av sin styrke når de vurderes sammen med de andre prediktorer. Regresjonsanalysen viser videre at de prediktorer som eksisterer allerede før opptaket gir en god prediksjon av fremtidige prestasjoner. De benyttes til å gjennomføre en pre-seleksjon, og avgjør hvem som skal kalles inn til selve opptaket. Det kan derfor tolkes dit hen at valget av hvem som skal kalles inn til opptak gjennomføres med bakgrunn i gode prediktorer. Det tilføres imidlertid ytterligere forklaring av total varians i løpet av opptaket, selv om to av prediktorer ikke fremstår som unike bidragsytere. Skoleprognosen fra intervjuet mister noe av sin styrke når den settes inn i en analyse med blant annet skolepoeng fra videregående. Det er sammenfallende med funnene i regresjonsanalysen mot akademisk prestasjon i Tabell II. Lederprognosen har både gjennom korrelasjonsanalysen og de andre regresjonsanalysene fremstått med relativt lav prediktiv validitet. Samlet sett viser regresjonsanalysen for total prestasjon etter skoleåret ved befalsskolen at de prediktorer som fremstår med høyest prediktiv validitet er skolepoeng videregående, AE, samt offisersvurdering FOS (vurderingssenter).

Alt i alt kan det se ut som de to prediktorer som eksisterer før opptaket, skolepoeng fra videregående og AE, gir et godt grunnlag for seleksjon til befalsskolen. Det gjelder først og fremst prediksjon av akademiske evner og total prestasjon. De kan se ut til å predikere dårligere opp mot lederpotensiale. Selv om de to prediktorer gir et godt grunnlag, viser regresjonsanalysene i denne studien at det tilføres bedre prediksjon av fremtidige prestasjoner underveis i opptaket. I første omgang gjelder det spesielt offisersvurdering FOS fra feltøvelsen, men også skoleprognosen fra intervjuet tilfører noe. Flere studier underbygger at det er

kombinasjonen av flere prediktorer som kvalitetshever seleksjonen, og at de således er gjensidig avhengig av hverandre for å oppnå gode resultat (Melchers & Annen, 2010; Schmidt & Hunter, 1998). Utfordringen blir dermed å finne den beste kombinasjonen av prediktorer gitt de kriterier en velger å evaluere mot.

Regresjonsanalysene viser videre at akademiske evner og total prestasjon predikeres best. Total forklart varians av lederskap viser til noe lavere verdier. Dersom en tar utgangspunkt i kandidatenes erfaringsgrunnlag er kanskje det ikke spesielt overraskende. De har gått flere år på skole og det finnes således et godt grunnlag for å kunne vurdere potensiale og prestasjoner innen akademiske evner. Når det gjelder lederskap har de aller fleste ingen erfaring, både hva gjelder teoretisk plattform eller praksis innen det faget. Det kan derfor synes som om det er vanskeligere å kunne forutsi lederpotensiale, enn akademiske evner. På en annen side viser både korrelasjonsanalysen og regresjonsanalysene at prediktoren offisersvurdering FOS fremstår som en god prediktor av nettopp lederskap.

Allikevel er det fortsatt mange andre forhold som forklarer varians både for akademiske evner og lederskap. Utfordringen videre blir dermed å forbedre seleksjonssystemet slik at forklart varians blir noe høyere.

Etter å ha diskutert funnene for korrelasjons- og regresjonsanalysen, følger nå en gjennomgang av styrker og svakheter ved studien.

### **6.3 Styrker og svakheter ved studien**

En av styrkene ved oppgaven er tallmaterialet. Det har ikke vært noen direkte kontakt mellom verken kandidatene eller intervjuoffiserer og meg som forsker. På den måten har ikke studien eller forskeren påvirket utfall eller vurderinger utført under opptaket (Jacobsen, 2003, s. 216-218). På en annen side kunne samtaler med intervjuoffiserer i etterkant av opptaket tilført større forståelse av hvordan de tenkte underveis i opptaket. Spesielt kunne det vært nyttig i forbindelse med problematikken rundt setting av skole- og lederprognosen.

En annen styrke ligger i at det nettopp er tallmateriale som benyttes i studien. Det foreligger ingen tolkning eller vurderinger på de enkelte tallverdier ut over det å gjennomføre kontroller i form av deskriptiv statistikk som normalfordeling, multikolinearitet, standardavvik etc. På den måten er ikke tallmaterialet utsatt for påvirkning av selve studien. På en annen måte er derimot tolkning av resultat og i tillegg oppbygging av multipl hierarkisk regresjonsanalyse et resultat av meg som forsker. En annen inndeling og rekkefølge i regresjonsanalysen ville ha gitt andre resultater. Valgt metode og oppbygging er derfor forsøkt bygget opp på en mest mulig naturlig måte i forhold til faktisk rekkefølge og gjennomføring av selve opptaket.

Noe av svakheten i denne studien er at en stor del av det totale utvalget er sendt hjem og kan således ikke sjekkes opp mot de kriterier som er etablert. Dersom disse kandidatene hadde levert resultater i form av eksamener og andre vurderinger ville en kunne oppnådd høyere korrelasjoner enn hva som vises til i denne oppgaven. Problemet omtales som *restriction of range*, og er behandlet tidligere i studien. Det er noe av utfordringen i forhold til å vurdere prediktiv validitet i forbindelse med seleksjon.

En annen svakhet kan være valg av kriterier. For å kunne avdekke prediktiv validitet for akademisk prestasjon er det grunn til å mene at resultat fra skoleeksamener i teorifag er et godt kriterium. Som nevnt i diskusjonsdelen kan det sies å være en større utfordring knyttet til offisersvurdering HBS som kriterium innen lederskap. Kriteriet benyttet i oppgaven er en vurdering etter ett års skolegang, og lederskap må kun sies å være et modningsfag. Anvendelse og utvikling av lederskap er en kontinuerlig og livslang prosess. Det er mulig at en vurdering av lederskap foretatt noe senere i karrieren hadde kunne gitt et bedre kriterium. På en annen side vil ikke alle som utdannes gjennom befalsskolen fortsette i Forsvaret, og således ville utvalget kunne bli enda mindre.

I denne studien sammenlignes og diskuteres funnene opp mot tidligere studier som er relevante og sammenlignbare. I noen tilfeller kan det være utfordrende å avdekke hvordan de enkelte prediktorer gjennomføres i de enkelte studier. På samme måte gjelder det for kriterier. Det er ikke alltid like lett å se hvordan de ulike prestasjoner faktisk måles. På den måten kan sammenligninger av studier medføre en noe upresis diskusjon. Det er i denne studien forsøkt marginalisert ved å velge ut studier som kan se ut til å være mest mulig lik, både i forhold til prediktor og kriterium.

## 7 Konklusjon

Denne studien tar utgangspunkt i et overordnet spørsmål rundt den prediktive validiteten til de prediktorer som benyttes ved årlig opptak og seleksjon til befalskolen. Studien bekrefter at det er sammenheng mellom prediktorer og kriterier ved opptaket. Ikke overraskende varierer denne sammenhengen for de ulike prediktorer, avhengig av hvilke kriterier som de måles opp mot. De to prediktorer som eksisterer allerede før selve opptaket, skolepoeng videregående og AE, viser til relativt god sammenheng med skolerresultat ved befalsskolens slutt. Mer presist er de høyest korrelert med akademisk prestasjoner, mens de viser noe lavere korrelasjoner med lederpotensiale. De samme prediktorer viser i tillegg til god sammenheng med total prestasjon etter endt skoleår. Ut over det viser regresjonsanalysen at de to prediktorer forklarer mye av variansen for akademisk prestasjon, sett opp mot skoleprognosen som er den andre prediktoren som skal forutsi evnen til å prestere teoretisk. Her må det tilføyes at dersom en hadde snudd på rekkefølgen i regresjonsanalysen, hadde skoleprognosen vist til god forklaring av varians. Nå er det imidlertid slik at både skolepoeng vgs. og AE eksisterer allerede før opptaket, slik at det er naturlig at de legges inn først. Det samme bildet avtegner seg i forhold til å predikere opp mot total prestasjon ved befalsskolen. Skolepoeng videregående og AE forklarer mest av variansen for både akademisk og total prestasjon, når de settes inn i en regresjonsmodell med de prediktorer som etableres under opptaket. På en side kan det antyde at de to prediktorer som eksisterer allerede før opptaket bør gis en mer selvstendig rolle sett opp mot det endelige opptaket. I dag teller de inn kun i forhold til å bli tatt opp som aspirant og videre deltagelse på feltøvelsen (vurderingssentret). Allikevel er det fullt mulig å se på resultatet som en støtte for dagens modell på den måten at de prediktorer som benyttes for å velge ut kandidater som skal innkalles til selve opptaket, baserer seg på gode måleparametre.

Videre viser studien til at det er sammenheng mellom de to prognosene som settes på bakgrunn av intervjuet og senere prestasjoner. Skoleprognosen viser til en klart høyere sammenheng med akademisk prestasjon enn lederprognosen gjør med lederpotensiale. Sammenhengen innbyrdes mellom de to prognosene kan kanskje antyde at den ene prognosen i for stor grad påvirkes av den andre. Skoleprognosen viser en god sammenheng med både akademisk prestasjon, men også en god sammenheng med total prestasjon. På den måten viser prognosen til en relativt høy prediktiv validitet. På en annen side viser regresjonsanalysen at prognosen kanskje ikke tilfører så mye ekstra forklaringskraft hva gjelder variansen for akademisk og total prestasjon. Samlet sett har derfor skoleprognosen god korrelasjon med det den skal predikere, men samtidig tilfører den kanskje ikke så mye som ønsket hva gjelder

forklaring av varians. Allikevel tilfører den noe, og kan derfor sies å fortjene en selvstendig rolle i det endelige opptaket.

Lederprognosen viser til tilnærmet lik sammenheng med alle de tre kriteriene i studien. Noe høyere sammenheng med total prestasjon enn med lederpotensiale. Forskjellene er imidlertid marginale. Det må kunne sies å være noe overraskende sett opp mot hensikten med selve prognosen. På en annen side kan forklaringen ligge i hvordan prognosen etableres i selve intervjuet. Funnene i studien viser til høy korrelasjon mellom skole- og lederprognosen, og således kan det se ut til at den ene påvirkes av den andre i for stor grad. Skoleprognosen starter med konkrete resultat i form av skolepoeng og AE, og er således tuftet på noe håndfast. Lederprognosen derimot, starter med mer blanke ark. Samlet sett kan det derfor antydes at vurdering av lederprognosen bør utvikles videre og gis et særskilt fokus både i opplæring av intervjuoffiserer, samt gjennomføring av selve intervjuet. Videre underbygges den noe svake korrelasjonen gjennom regresjonsanalysene. Lederprognosen viser til relativt lav forklaring av variansen for lederpotensiale. Det meste av forklaringen av variansen for lederpotensiale utgjøres av offisersvurdering FOS.

Når det gjelder sammenhengen mellom offisersvurdering FOS og senere prestasjoner viser studien også til en relativt god sammenheng. Vurderingssenter eller feltøvelsen med de vurderinger som gjøres kan derfor sies å ha en god sammenheng med prestasjoner innen lederskap etter endt skoleår. I tillegg viser denne prediktoren til relativt god sammenheng med total prestasjon. Den viser også til relativt god prediktiv validitet når det kommer til predikere lederpotensiale ved endt befalsskole. Allikevel kan det stilles spørsmål rundt lengde og kostnader ved feltøvelsen som leder frem til denne prediktoren. Det bør fortsatt være gjenstand for fortløpende vurderinger fremover.

### **7.1 Akademiske implikasjoner**

Når det gjelder akademiske implikasjoner av denne studien fremstår kanskje leder- og skoleprognosen som de prediktorer som krever videre studier. For det første er prognosene relativt nye, og organisasjonen har relativt lite erfaring med dem. I tillegg er dette den første undersøkelsen som ser spesifikt på de to prognosene hva gjelder opptak og seleksjon til befalsskolen. På grunn av det over har det så langt vært et begrenset datagrunnlag tilgjengelig for å gjennomføre slike studier. Fra sommeren av vil det være et betydelig større utvalg som kan analyseres. Da vil det være ett kull til fra Hærens befalskole, men i tillegg vil det kunne

gjennomføres analyser for befalskolekull fra Sjø- og Luftforsvaret. Det vil kunne frembringe en bredere og dypere analyse av spesielt de to nye prediktorer.

Et annet aspekt er lederprognosen, spesielt i forhold til valg av kriterium. I diskusjonsdelen antydes det at valg av kriterium kan spille en vesentlig rolle. På sikt kan det være viktig å se på sammenhengen mellom lederprognosen fra opptaket som prediktor, opp mot et kriterium på lederskapsprestasjoner senere i karrieren. En slik studie vil kunne underbygge de funn som er gjort i denne studien, alternativt vise til andre sammenhenger.

## 7.2 Praktiske implikasjoner

Hvilke praktiske implikasjoner fremkommer av denne studien? Først og fremst utpeker lederprognosen seg som den prediktoren som kan sies å skåre dårligst i forhold til forventninger og prediktiv validitet. Som nevnt over bør det vurderes å sette et spesifikt fokus på denne prognosen i forberedelsene til neste opptak og seleksjon. I første omgang vil problematikken rundt påvirkning fra skoleprognosen være det viktigste å ta tak i. Bevissthet rundt det fenomenet, samt et enda strammere rammeverk for hva en ser etter og vurderer i intervjuet kan være en mulig løsning for å skape en høyere validitet og selvstendighet for denne prediktoren. Samtidig er det viktig å ikke ta forhastede konklusjoner basert på denne ene studien. Det er første gang den aktuelle prognosen benyttes som prediktor, og det er første gang den statistisk analyseres på et relativt beskjedent tilgjengelig utvalg.

En annen praktisk implikasjon kan være bruk av fysisk test ved FOS som selvstendig prediktor i det endelige opptaket til elev. Prediktoren viser til lav prediktiv validitet med alle kriterier i denne studien. Det betyr ikke at den er uviktig som måleparameter i en seleksjonsprosess, men det kan stilles spørsmål ved den selvstendige rollen den har i det endelige opptaket. I dag vektes den med 10 % ved opptaket. Det kan kanskje vurderes om den bør bli en test som opererer med bestått/ikke bestått under opptaksuken. Det frigjør plass til enten å gi de andre prediktorer en høyere vekting, eller skape rom for andre dimensjoner. Studien peker på en mer sentral rolle for skolepoeng videregående skole og AE, som et eksempel.

En seleksjonsprosess som FOS er som tidligere nevnt en stor kostnadsdriver. Vurderinger rundt kost/nytte har også hatt sin sentrale plass i tidligere forskning. Det er imidlertid noe diskusjon rundt hvordan slike analyser og vurderinger best bør gjennomføres (Schmidt et al., 1992, s. 651-654). Likevel er det enkelte som har vurdert og til og med verdsatt i kroner den nytten enkelte seleksjonsmetoder medfører (Jones, 1989, s. 73). Uavhengig om nytten beregnes i antall kroner eller om den verdivurderes i form av økt produksjon, tyder flere av tidligere forskningsprosjekter på at en god seleksjon med prediktorer som viser til høy prediktiv validitet

gir en positiv nytte sett opp mot kostnadene (Schmidt & Hunter, 1998, s. 273). Det understøttes videre av forskning som har sett på militær seleksjon spesifikt over en lengre tidsperiode (Campbell & Knapp, 2010, s. 882). På denne måten blir videre forskning og utvikling av Felles Opptak og Seleksjon til befalsskolen viktig også videre fremover. Kostnaden ved feil utvelgelse av kandidater kan raskt komme opp i store summer. Utfordringen blir derfor å finne den optimale balansen mellom de ressurser som benyttes sett opp mot den totale nytten i ettertid. Vurderinger rundt dette gjelder først og fremst lengde og innhold for feltøvelsen, men bør også vurderes i forhold til hele seleksjons- og opptaksprosessen.

Alt i alt viser denne studien at de fleste prediktorer som benyttes ved Felles Opptak og Seleksjon ved befalsskolen har en høy prediktiv validitet. Resultatene varierer selvfølgelig ut fra hvilke kriterier som det evalueres opp mot. Uavhengig av resultatene i denne studien vil det være viktig å kontinuerlig evaluere seleksjonsprosessen slik at en hele tiden strever etter en kvalitativt best mulig seleksjon. Noen av de kandidatene som velges ut kan om kort tid befinne seg i tilsvarende situasjon som beskrevet i eksemplet fra Afghanistan i starten av studien.



## Referanser

- Alf, E. F., Neumann, I., & Mattson, J. D. (1988). *Revision of the United States Naval Academy selection composite* (NPRDC TN 88-61). San Diego: Navy personnel research and development center.
- Arthur, W. J., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology*, 56(1), 125-154.
- Bartram, D. (2001). Guidelines for test users: A review of national and international initiatives. *European Journal of Psychological Assessment*, 17, 173-186.
- Bentzrød, S. B. (2011, 17.mar). Mister offiserene sine når de er 35 år. *Aftenposten*. Hentet fra <http://www.aftenposten.no/nyheter/iriks/article4064325.ece>
- Borack, J. I. (1994). Estimating predictive validity when range restriction due to selection and attrition is present. *Military Psychology*, 6(3), 193-204.
- Braun, P., Wiegand, D., & Aschenbrenner, H. (1991). The assessment of complex skills and of personality characteristics in military services. I R. Gal & D. Mangelsdorff (Red.), *Handbook of Military Psychology*: Chichester: John Wiley & Sons Ltd.
- Campbell, J. P., & Knapp, D. J. (2010). Project A, 12 years of R & D. I J. L. Farr & N. T. Tippins (Red.), *Handbook of Employee Selection* (s. 865-886). New York: Routledge Taylor & Francis Group.
- Dakin, S., & Armstrong, J. S. (1989). Predicting job performance: A comparison of expert opinion and research findings. *International Journal of Forecasting*, 5(2), 187-194.
- Dobson, P., & Williams, A. (1989). The validation of the selection of male British Army officers. *Journal of Occupational Psychology*, 62, 313-325.
- Farr, J. L., & Tippins, N. T. (2010). Handbook of employee selection. An introduction and overview. I J. L. Farr & N. T. Tippins (Red.), *Handbook of employee selection* (s. 1-6). New York: Routledge Taylor & Francis Group.
- FFOD. (2007). *Forsvarets fellesoperative doktrine*. Oslo: Forsvarsstaben.
- Forsvarets skolesenter. (2007). *Tjenestereglement for Forsvaret kl 4 - Utdanning av befal*. Oslo: Forsvaret
- Forsvarsdepartementet. (2008). *St.prp.nr.48 (2007-2008):Et forsvar til vern om Norges sikkerhet, interesser og verdier*. Oslo: Regjeringen.
- Gebhardt, D. L., & Baker, T. A. (2010). Physical performance tests. I J. L. Farr & N. T. Tippins (Red.), *Handbook of employee selection* (s. 277-298). New York: Routledge Taylor & Francis Group.

- Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology*, 18(2), 135-164.
- Hansen, I. (2007). *Bidrag til Psykologitjenestens historie i Forsvaret fra 1946-2006*. Oslo: Forsvarets Institutt for Ledelse.
- Hardison, C. M., Sims, C. S., & Wong, E. C. (2010). *The air force officer qualifying test. Validity, Fairness and Bias* (RAND Project Air Force).
- Hermelin, E., Lievens, F., & Robertson, I. T. (2007). The validity of assessment centres for the prediction of supervisory performance ratings: A meta-analysis. *International Journal of Selection and Assessment*, 15(4), 405-411.
- Hough, L., & Dilchert, S. (2010). Personality - Its measurement and validity for employee selection. I J. L. Farr & N. T. Tippins (Red.), *Handbook of employee selection* (s. 299-319). New York: Routledge Taylor & Francis Group.
- Hulsheger, U. R., Maier, G. W., & Stumpp, T. (2007). Validity of general mental ability for the prediction of job performance and training success in Germany: A meta-analysis. *International Journal of Selection and Assessment*, 15(1).
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological bulletin*, 96(1), 72-98.
- Hunter, J. E., & Schmidt, F. L. (2003). *Methods of meta-analysis: Correcting error and bias in research findings*. Beverly Hills: CA: Sage.
- Hystad, S. W., Eid, J., Laberg, J. C., & Bartone, P. T. (2011). Psychological hardiness predicts admission into Norwegian military officer schools. *Military Psychology*, 23, 381-389.
- International Task Force on Assessment Center. (2000). *Guidelines and Ethical considerations for Assessment Center operations*. Paper presented at the 28th International Congress on Assessment Center methods, San Francisco.
- Jacobsen, D. I. (2003). *Hvordan gjennomføre undersøkelser? Innføring i samfunnsvitenskapelig metode*. Kristiansand: Høyskoleforlaget AS.
- Johannessen, A. (2003). *Intro til SPSS*. Oslo: Abstrakt forlag.
- Jones, A. (1989). The contribution of psychologist to military officer selection. I R. Gal & D. Mangelsdorff (Red.), *Handbook of Military Psychology*. Chichester: John Wiley & Sons Ltd.
- Judge, T. A., Colbert, A. E., & Ilies, R. (2004). Intelligence and leadership: A quantitative review and test of theoretical propositions. *Journal of Applied Psychology*, 89(3), 542-552.

- Kahneman, D. (2011, 19 oktober). Don't blink! The hazards of confidence. *The New York Times Magazine*. Hentet fra <http://www.nytimes.com/2011/10/23/dont-blink-the-hazards-of-confidence.html?>
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.
- Knapp, D. J., & Tremble, T. R. (2007). Development of experimental Army enlisted personnel selection and classification test and job performance criteria. *Technical Report 1168*. Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- Lance, C. E., Woehr, D. J., & Meade, A. W. (2005). A Monte Carlo investigation of assessment center construct validity models. *Paper presented at the 20th Annual conference of the Society for Industrial and Organizational Psychology*.
- Marcinik, E. J., Hyde, D. E., & Taylor, W. F. (1995). The relationship between the U.S. Navy fleet diver physical screening test and job task performance. *Aviation, Space and Environmental Medicine*, 66(4), 320-324.
- Martinussen, M., & Hunter, D. (2008). *Luftfartspsykologi*. Bergen: Fagbokforlaget.
- Martinussen, M., & Torjussen, T. M. (2004). Initial validation of a computer-based assessment battery for pilot selection in the Norwegian Air Force. *Human Factors and Aerospace Safety*, 4, 225-232.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of applied psychology*, 79(4), 599-616.
- Melchers, K. G., & Annen, A. (2010). Officer Selection for the Swiss Armed Forces. *Swiss Journal of Psychology*, 69(2), 105-115.
- Morgeson, F. P., & Campion, M. A. (2000). Accuracy in job analysis: Toward an inference-based model. *Journal of Organizational Behavior*, 21, 819-827.
- Murphy, K. R. (2008). Explaining the weak relationship between job performance and ratings of job performance. *Industrial and organizational psychology*, 1, 148-160.
- Nathan, B. R., & Tippins, N. T. (1990). The consequences of Halo "Error" in performance ratings: A field study of the moderating effect of Halo on test validation results. *Journal of Applied Psychology*, 75(3), 290-296.
- Norsk samfunnsvitenskapelig datatjeneste AS. ( 16 mar 2012). *Tilbakemelding på melding om behandling av personopplysninger*. Bergen: NSD
- Pallant, J. (2010). *SPSS Survival manual. A step by step guide to data analysis using SPSS* (4. utg.). Berkshire: Mc Graw Hill.

- Park, H. M. (2010). Hypothesis testing and statistical power of a test. Working paper. *The University Information Technology Services (UITS) Center for Statistical and Mathematical Computing, Indiana University.*
- Petrides, K. V., Weinstein, Y., Chou, J., Furnham, A., & Swami, V. (2010). An investigation into assessment centre validity, fairness and selection drivers. *Australian Journal of Psychology*, 62(4).
- Pursell, E. D., Champion, M. A., & Gaylord, S. R. (1980). Structured interviewing: Avoiding selection problems. *Personnel Journal*, 59(11), 907-912.
- Rayson, M., Holliman, D., & Belyavin, A. (2000). Development of physical selection procedures for the British Army. Phase 2: Relationship between physical performance tests and criterion tasks. *Ergonomics*, 43(1), 73-105.
- Ringdal, K. (2009). *Enhet og mangfold. Samfunnsvitenskapelig forskning og kvantitativ metode* (2. utg.). Bergen: Fagbokforlaget.
- Rones, N., & Fasting, K. (2011). *Befalsskolestudien - Felles opptak og seleksjon 2010*. Oslo: Forsvarets Høgskole Norges Idrettshøgskole/Forsvarets institutt.
- Russell, C. J., & Domm, D. R. (1995). Two field tests of an explanation of assessment centre validity. *Journal of Occupational and Organizational Psychology*, 68, 25-47.
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., De Fruyt, F., & Rolland, J. P. (2003). A meta-analytic study of General Mental Ability validity for different occupations in the European community. *Journal of Applied Psychology*, 88(6), 1068-1081.
- Salgado, J. F., Anderson, N. R., & Hulsheger, U. R. (2010). Employee selection in Europe. Psychotechnics and the forgotten history of modern scientific employee selection. I J. L. Farr & N. T. Tippins (Red.), *Handbook of employee selection* (s. 921-941). New York: Routledge Taylor & Francis Group.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262-274.
- Schmidt, F. L., Ones, D. S., & Hunter, J. E. (1992). Personnel selection. *Annu. Rev. Psychol.*, 43, 627-670.
- Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, 32, 349-362.
- Trost, J., & Jeremiassen, L. (2010). *Interview i praksis*. København: Hans Reitzels Forlag.

- Wiesner, W. H., & Cronshaw, S. F. (1988). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. *Journal of Occupational Psychology*, 61, 275-290.
- Withman, D. S. (2009). Emotional intelligence and leadership in organization: A meta-analytic test of process mechanisms. *FIU Electronic These and Dissertations*. Hentet fra <http://digitalcommons.fiu.edu/etd/113>
- Woehr, D. J., & Arthur, W. J. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management*, 29(2), 231-258.
- Zickar, M. J., Cortina, J. M., & Carter, N. T. (2010). Evaluation of measures: Sources of error, sufficiency and contamination. I J. L. Farr & N. T. Tippins (Red.), *Handbook of Employee selection* (s. 399-415). New York: Routledge Taylor & Francis Group.